Author:
Avila Castro, Alberto

Title:
Peptide Morse code

/-peptide -sheets for information storage and transmission

Peptide Morse code: α/β-peptide β-sheets for information storage and transmission

Alberto Avila Castro

Supervisor:

Prof. Anthony P. Davis

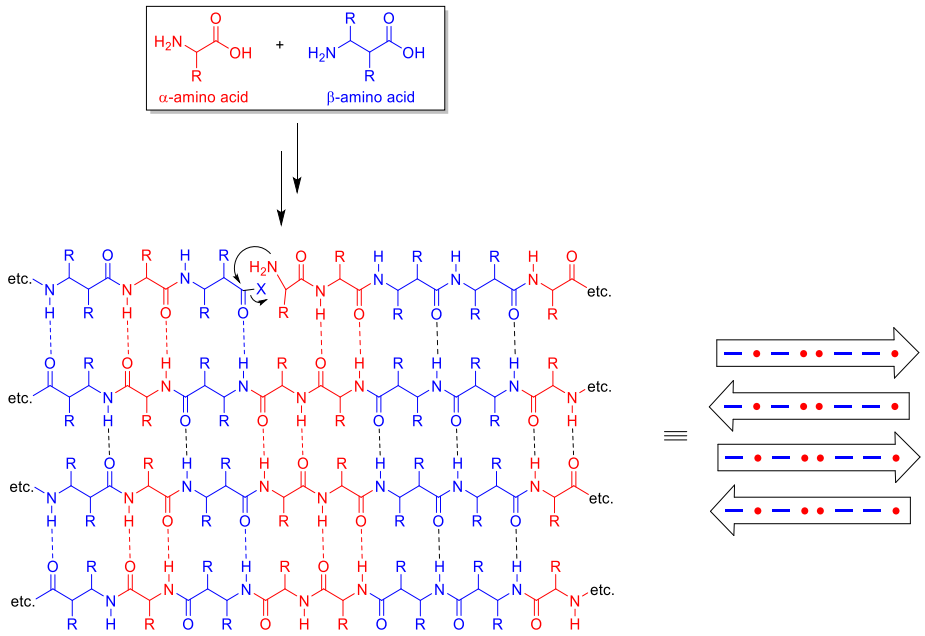
March 2020

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

Word count: 39, 895

Abstract

The origin of life remains a most relevant scientific problem. It is clear that present-day life results from Darwinian evolution. However, in a prebiotic Earth, species (presumably molecules) capable of "informational self-replication" were necessary for evolution to be triggered. In this context, we propose a system based on α/β-peptide hybrid β-sheets. In these molecules, information can be encoded in the sequence of α- and β-amino acids. Aggregation into β-sheets should occur sequence-selectively and the aggregate can then act as a template for peptide self-replication. The use of long/short components to store and transfer information suggests the term "peptide Morse code" (PMC) for the system.



To prove this concept, we took two approaches: First, to study the sequence-selectivity of β-sheet formation, we used a series of short decamer α/β-peptide hybrids containing a β-turn segment. It was shown through NMR and CD analyses that molecules with matching α/β-residues in their β-strands would fold into stable β-hairpins in organic media, due to the formation of an intramolecular β-sheet, whereas, there was structural evidence for the absence of such stable intramolecular β-sheets in molecules bearing strands with mismatching α/β-residues. Secondly, to evaluate sequence-selective self-replication in α/β-peptide hybrids, we built a water-soluble amphiphilic PMC variant of a literature replicator and an alternative hydrophobic system. Unfortunately, autocatalysis has remained elusive in both peptide systems. Alternatively, evidence for the intramolecular templation of new covalent bonds arising from α/β-sequence recognition in the strands of a β-hairpin was obtained.

This work has provided the first evidence for both, the α/β-sequence-selective assembly of β-sheet structures and the template-directed formation of covalent bonds by these structures.

Acknowledgments

I would like to thank Professor Anthony Davis for giving me the opportunity to work in this project and for his guidance and support throughout.

Thanks to the members of the Davis group, past and present: Rob Tromans, Miriam Wilson, Sergio Benitez, Claire Webster, Danny Van Eker, Laurent Chabanne, Guillem Peñuelas, Chris Dias, Tom Carter, Nisar Ahmed, Jinqiao Dong, Kelly Chu, Danny Burke, Tom Wieczorek, Mike Orchard, Ash Griffith, Patrick Stewart, Amir Shermila Kaspar, Bahar Shirinfar, Germinal Magro, Conor Lanphere, Vito Zhao, Martin Doll, Eva González Freire, Eléa Wallisky, Marion Kieffer, Soumen Samanta, Daniel Austin, Federica Balduzzi, Al Flint, Nick Wragg, Alice Quaife, Sophie Remillard, Joe Mclaughlin, Millie Giles and all other visitors. Thanks for your patience and for being so helpful. Dr Nisar Ahmed is acknowledged for introducing me to peptide chemistry when I first joined the group. I am especially grateful to members of the "Origin of Life team": Dr Nisar Ahmed, Dr Jinqiao Dong, Kelly Chu, Danny Burke and Tom Wieczorek. Their work has contributed significantly to the project, including parts of this thesis. It has been extremely rewarding to work with you.

Thanks to all technical staff and collaborators at Bristol for their assistance with this project, especially Dr Matt Crump, Paul Lawrence and Dr Chris Arthur.

Thanks are necessary for the Bristol Chemical Synthesis Centre for Doctoral Training for providing the opportunity and support, Emma Rose, Mar Ruíz Molina, Kevin Booker-Milburn and Laura Chavda. I also extend my thanks to everybody in the 2015 CDT cohort for their support.

I would also like to thank the EPSRC, CONACYT and Tony for funding this PhD project.

Additional thanks to Dr Javier García Cárceles and Claire Webster for proof reading parts of this thesis.

I would like to thank my friends, especially the ones I have made during my time in Bristol. Diego, Sergio, Javi, Claire, Danny, Inma and Javi Ramos thank you for your patience, help, and friendship.

Finally, I would like to thank my family, especially my parents Oralia and Alberto, my brother Héctor, my uncles Zacarias and Evelia, and my cousin Polo for their patience, love, and support.

Dedico este trabajo a mis abuelos Graciela, Héctor y Ramona, y a mi tía Evelia.

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....Alberto Avila Castro..... DATE:…20/03/2020…

Table of contents

Table of figures

Table of schemes

Table of tables

Abbreviations

ABA	4-Acetamidobenzoic acid
ACN	Acetonitrile
AcOH	Acetic Acid
Bn	Benzyl
°C	Celsius degrees
CD	Circular dichroism
conc.	Concentration
Da	Daltons
DCM	Dichloromethane
DIPEA	*N*,*N*-Diisopropylethylamine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Fmoc	Fluorenylmethyloxycarbonyl
GdmCl	Guanidinium chloride
h	Hours
HATU	Hexafluorophosphate azabenzotriazole tetramethyl uronium
HSQC	Heteronuclear single quantum coherence spectroscopy
ISR	Informational self-replication
LUCA	Last universal common ancestor
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
Me	Methyl
MESNA	Sodium 2-mercaptoethanesulfonate
min	Minutes
mL	Millilitres

mM	Millimolar
MPAA	(4-Carboxymethyl)thiophenol
MRE	Mean residue ellipticity
µm	Micrometres
NCL	Native Chemical Ligation
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OoL	Origin of life
HOSu	*N*-Hydroxysuccinimide
PEG	Polyethylene glycol
PyBop	Benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate
ppb	Parts per billion
RNA	Ribonucleic acid
ROESY	Rotating frame Overhauser effect spectroscopy
HPLC	High-Performance Liquid Chromatography
rt	Room temperature
SPPS	Solid-phase peptide synthesis
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TIPS	Triisopropylsilane
TOCSY	Total correlation spectroscopy
t_R	Retention time
XNA	Xenonucleic acid

The amino acids and their three- and one-letter abbreviations described in this thesis are as follows:

Alanine	A	Ala
Arginine	R	Arg
β-Alanine	-	β-Ala
Cysteine	C	Cys
Glutamic acid	E	Glu
β-Homoglutamic acid	-	β^3hGlu
Glutamine	Q	Gln
Glycine	G	Gly
Leucine	L	Leu
β-Homoleucine	-	β^3hLeu
Lysine	K	Lys
β-Homolysine	-	β^3hLys
Methionine	M	Met
Phenylalanine	F	Phe
β-Homophenylalanine	-	β^3hPhe
Proline	P	Pro
D-Proline	^{D}P	^{D}Pro
Threonine	T	Thr
Tyrosine	Y	Tyr
Valine	V	Val
β-Homovaline	-	β^3hVal

Chapter 1. Introduction

1.1. The study of the origin of life

The origin of life (OoL) is one of the most relevant unsolved problems for modern science,[1–3] and perhaps the most important with chemistry at its core.[4,5] How did inanimate matter transition to the living state on early Earth? (Figure 1.1) In principle, the problem can be studied from two opposite approaches: chemistry up and biology down.[6] On the one hand, prebiotic chemists are dedicated to elucidate plausible routes to life's building blocks, within the range of environmental scenarios and starting from precursors presumably available in prebiotic Earth.[7–9] On the other, biology can only go so far down to a putative last universal common ancestor (LUCA)[10,11], a primitive cell with most of the central biochemical machinery of present-day biology already installed.[4,12] However, elucidating the mechanism (or mechanisms) responsible for the assembly of the very complex and specific set of molecules that constitute extant life from the chaos of the "primordial soup" around 4.0 billion years ago remains a difficult and puzzling challenge.[13,14]

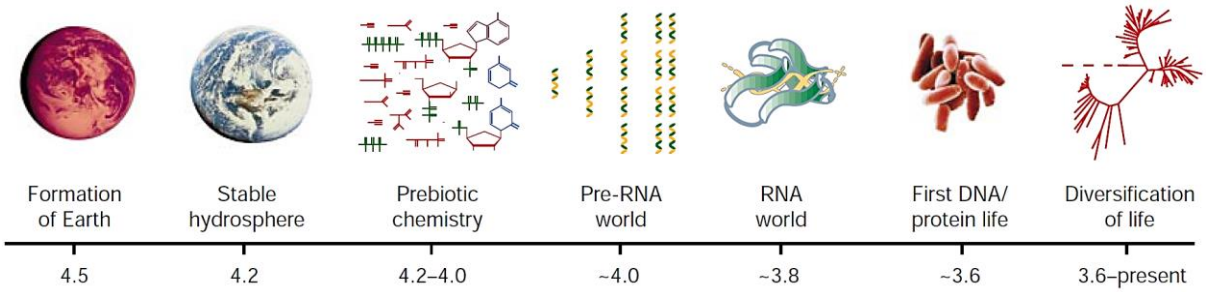


Figure 1.1: Timeline of events related to the emergence of life on Earth, with approximate dates in billions of years before the present (image taken from ref. 18).

1.2. Evolution and informational self-replication

A possible answer to this mystery is provided by the concept of evolution by natural selection, whereby entities competing for limited resources are preserved in preference to others in their local environment on the basis of intrinsic favourable characteristics.[15] Before living cells, this process would have relied on information-carrying molecular systems capable of self-replication (the emergence of life at its most basic level has been described as the emergence of information on a nanoscale[16] that can replicate[1]).

Chemically, self-replication consists on the ability of a molecule to catalyse the synthesis of accurate copies of itself from dispersed building blocks (autocatalysis). In its simplest form, a template molecule (**T**) recognises two components (**A** and **B**), binds to them reversibly and facilitates their coupling to

generate a new template product. This new molecule can then dissociate to provide a new template for further replications (Figure 1.2).[17]

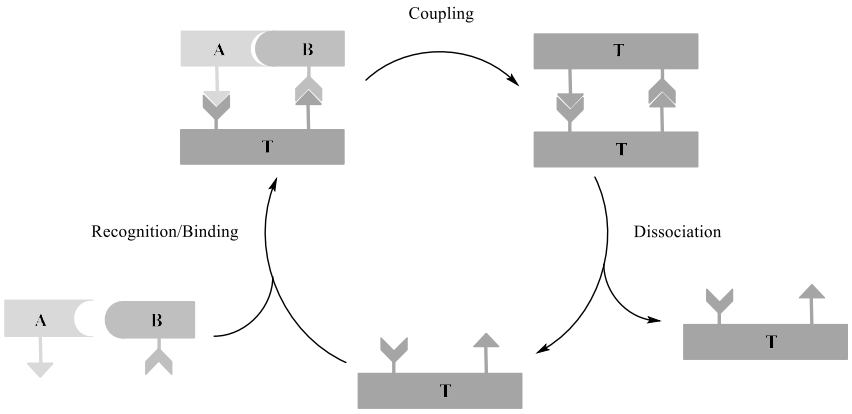


Figure 1.2: General representation of a "minimal" self-replicating chemical system.

"Informational self-replication" (ISR, Figure 1.3, left) requires molecules capable of great variety in their building block sequence (i.e. they must be able to store a lot of information, their genotype), so that each different sequence can interact with its environment in a specific way (phenotype). Each variant must be able to self-replicate accurately and undergo amplification, which means that a "parent" molecule will give rise to a large number of "progeny" molecules. Variants that are better fitted for this process will persist in their environment. These successful versions can then mutate (produce inaccurate progeny due to replication errors) and be subjected to further competence and selection (Figure 1.3, right).[5,15] In the early Earth context, this mechanism would have led to the creation of proto-life complexity from the abiotic clutter.[18]

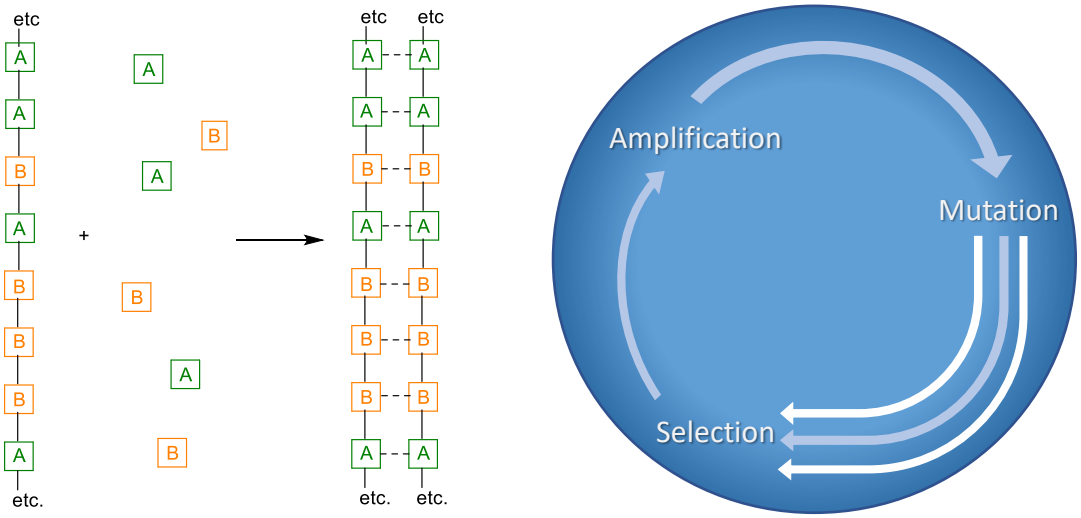


Figure 1.3: **Left.** Representation of an informational self-replicator. The molecule contains two components which can be incorporated in any order, effectively storing information. It can also template copies of itself from dispersed building blocks. **Right**. Darwinian evolution of informational systems. Amplification involves replication of parent molecules to produce large numbers of progeny molecules. Mutation introduces variation in the progeny. Selection chooses molecules that meet environmental constrains.

However, evolution by natural selection only provides a partial explanation to the OoL. How did this crucial process first become operational in prebiotic Earth? In modern life, ISR is a complex process carried out by a staggeringly sophisticated and efficient machinery based on DNA, RNA and proteins.[19,20] It is inconceivable that this machinery could have possibly emerged spontaneously from the "primordial soup". The self-replicator that triggered Darwinian evolution must have been simpler.[5]

1.3. The "RNA world"

There are strong arguments to support the hypothesis that modern DNA/protein-based life was preceded by a simpler life form based primarily on RNA. During this earlier stage, genetic information would have been stored in the sequence of RNA molecules and the phenotype derived from the metabolic enzyme-like properties of RNA (ribozymes).[18] Later on, DNA could have taken over genetic functions due to its greater stability, and polypeptides catalyst duties given their greater versatility.

Several observations regarding RNA's persistent role in modern biochemistry provide support to the "RNA world". Perhaps the most compelling is that a ribozyme lies at the heart of the ribosome, the protein-synthesising molecular machine.[20,21] Moreover, the developing field of *in vitro* evolution has shown that RNA can evolve in response to selection, proving that RNA could display Darwinian behaviour in the absence of cells.[15] Work carried out by Joyce and others[22–26] has demonstrated the

potential of this approach towards the development of ribozymes displaying outstanding RNA polymerase[i] capabilities.

This evidence makes the possibility of an "RNA world" difficult to ignore. However, the notion that RNA was the first replicator is more problematic.[5,18,25] As illustrated in Figure 1.4, RNA is a complicated molecule. Prebiotic synthesis of the component monomers (ribonucleotides) poses difficulties. Although quite a lot of encouraging work has been performed outlining routes in which this might have happened,[27–31] it still seems arguable whether high concentrations of the four canonical ribonucleotides could accumulate in the prebiotic world. Once formed, nucleotides must assemble into oligomers and, for ISR to occur, those oligomers should template the synthesis of complementary strands. However, in practice, template-directed replication of RNA (under controlled conditions and using canonical monomers only) has been difficult to achieve and is still under study.[4,23,25,32–36] Furthermore, there is the "strand inhibition problem", whereby transferring information from a template strand in the presence of its complementary strand is inhibited by the stability of the resulting duplex, which would also mean amplification is inhibited.

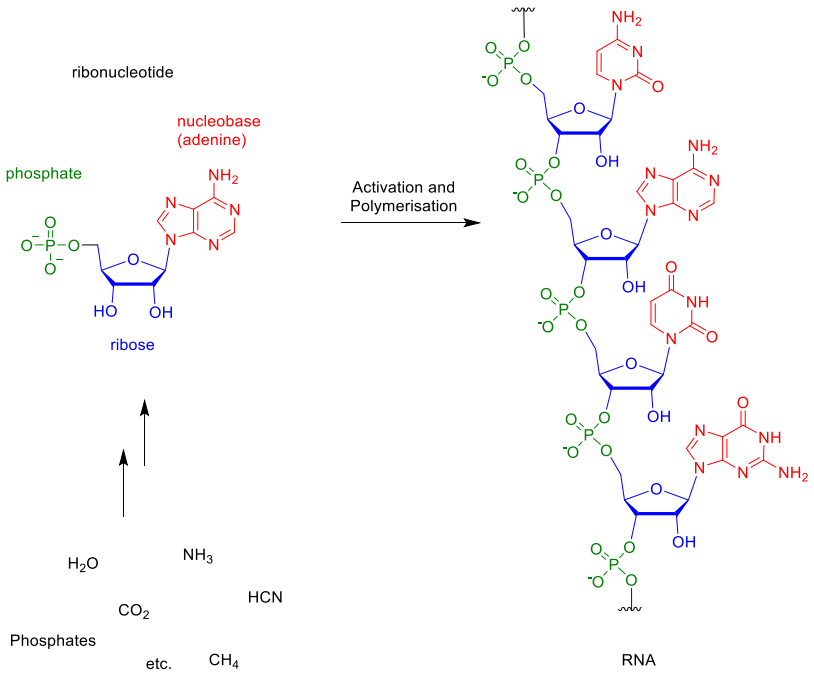


Figure 1.4: Structure of RNA and a simplified prebiotic route of synthesis.

Some of these issues may be solved by the catalytic action of a ribozyme which could promote (self-) replication.[23,25] However, such a long and complex molecule would need to spontaneously emerge from the prebiotic chaos, and although this is not impossible (it is noteworthy, that the synthesis of a true RNA replicase has proved elusive even under controlled laboratory conditions, see references 22-26),

[i] Enzymes capable of synthesising copies of RNA strands.

it leaves the door open for the consideration of alternative scenarios where RNA-based ISR was preceded in evolution by another kind of self-replicator which eventually led to the emergence of RNA.[5]

An approach often taken by prebiotic chemists involves designing informational polymers analogous to nucleic acids (the molecules known to display the most advanced informational self-replicating abilities), typically with simpler backbones or (including) non-canonical nucleobases, which makes their emergence more plausible in an abiotic context.[17,37–39] These molecules are generally termed xenonucleic acids (XNAs). However, even if this is an attractive hypothesis, it has not yet been unambiguously proved whether these nucleic acid analogues could have forerun and transitioned to RNA.[6,20,25,40]

Alternatively, it is possible that RNA-based life was preceded by a replicating polymer different to nucleic acids. Self-replication without Darwinian evolution has been demonstrated for certain systems. These have normally taken the form of "minimal self-replicators". Although the critical issue for these systems becomes whether they are capable of enough variety (able to encode a sufficient amount of information) and of replicating with sufficient fidelity[23] to provide the basis for evolution by natural selection.[18]

1.4. Peptide replicators

1.4.1. Helical peptides

If one seeks alternatives to nucleic acids, systems based on peptides are particularly attractive. The amino acid components of peptides are plausible prebiotic molecules, considerably simpler than nucleic acid monomers. The first self-replicating peptide was reported by Ghadiri.[41] It consisted of a 32-residue α-helix which could template its own synthesis from two smaller fragments through native chemical ligation (NCL, see section 2.1.2). The system is based on an α-helical coiled-coil tertiary structure where a hydrophobic core (Figure 1.5, left) serves as an interhelical recognition surface. The ligation site lies on the solvent exposed surface. Autocatalytic replication was established by monitoring the fragment ligation in the presence of various amounts of template (seeds). A marked increase in the initial rates of product formation was observed as the initial concentration of template was increased (Figure 1.5, right).

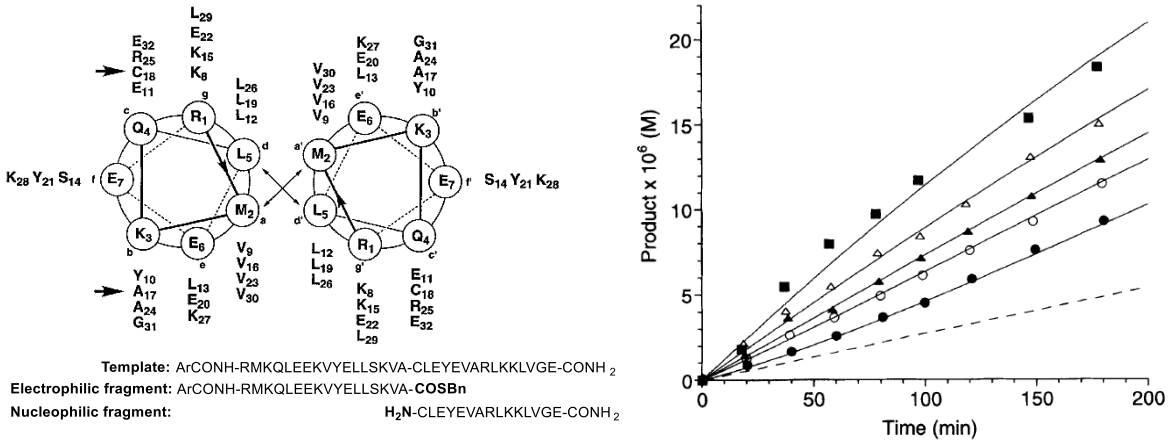


Figure 1.5: **Left**. Helical wheel diagram of the template peptide in the α-helical coiled-coil configuration emphasizing the ligation and the recognition sites (residues in the a and d positions). **Right**. Template production as a function of time for reaction mixtures containing different seeds of template: (filled circles) no added template, (open circles) 5 μM, (filled triangles) 10 μM, (open triangles) 20 μM and (filled squares) 40 μM (taken from ref. 41).

A closely-related peptide was reported by Chmielewski.[42,43] His system is pH sensitive, designed to form a coiled-coil under acidic conditions and to adopt a random coil conformation at physiological pH (Figure 1.6). At low pH the coupling between two fragments (**E1** and **E2**) to form the template (**E1E2**) proceeded *via* an autocatalytic template replication pathway. At neutral pH, however, the coupling proceeded in a noncatalytic manner. It is worth noting that these type of helical systems present some degree of product inhibition (due to unfavoured dimer dissociation).[44]

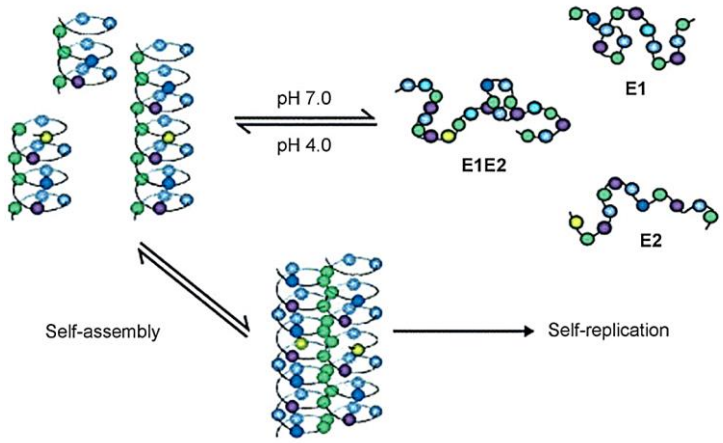


Figure 1.6: A pH-modulated self-replicating peptide, **E1E2** denotes peptide template, **E1** and **E2** denote the two peptide fragments (taken from ref. 43).

In an impressive subsequent report, Ghadiri demonstrated his replicator was sensitive to subtle changes in the polymer sequence (the modification of a single residue).[45] This system consisted of the original template peptide and three "mutant" variants; two of them, **T9A** and **T26A**, carrying a single amino

acid modification, and one, **T9/26A**, with two amino acid modifications (Figure 1.7, top). The mutations consisted of the substitution of a valine and/or leucine residues located in the hydrophobic core for alanine residues. The mutant peptides were shown to be produced in a non-catalytic manner from the corresponding nucleophilic and electrophilic fragments, but **T9A** and **T26A** were able to catalyse the formation of the native peptide **T** (Figure 1.7, catalytic cycles **II** and **III**, bottom). On the other hand, **T** was shown to be a selfish replicator since it did not catalyse the formation of any of the mutant variants, and the double mutant **T9/26A** was entirely non-catalytic. Due to these relationships, the native template (**T**) consistently dominated reaction mixtures. This was claimed as an "error correction mechanism".

This error correction was extended to the efficient amplification of homochiral products from racemic mixtures of peptide fragments through a chiroselective autocatalytic cycle. Again, it was also shown that structures bearing single residue stereochemical mutations could be discriminated against and presented impaired autocatalysis. Templates with only one L/D substitution promoted the production of the homochiral products only.[46] It has been postulated that this error correction mechanisms could qualify as the emergence of some form of primitive replication fidelity.[17,44] However, the *de novo* design of many variants of these molecules, each displaying similar behaviour would present a great challenge.

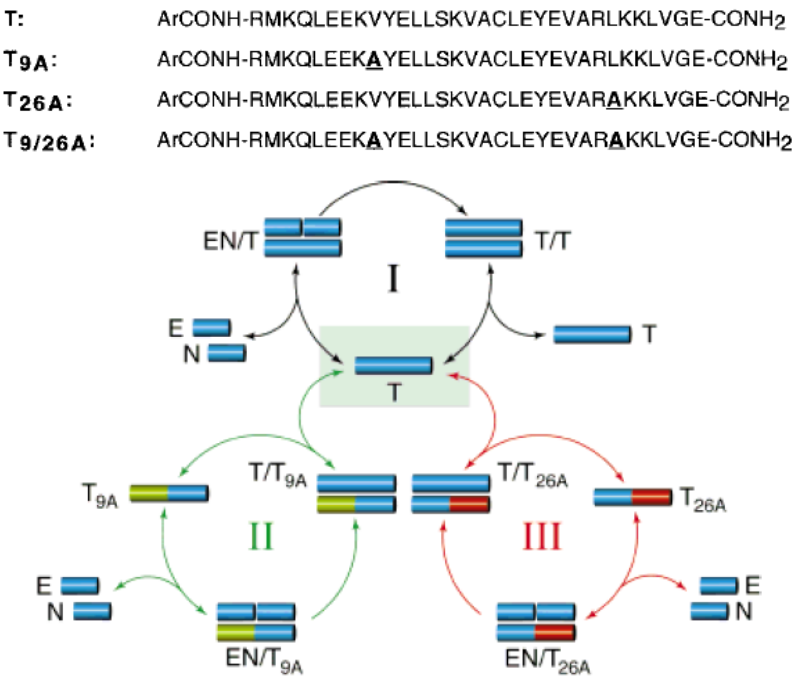


Figure 1.7: **Top**. Amino acid sequence of Ghadiri's templates with mutation sites highlighted. **Bottom**. Diagram summarizing the network of catalytic activity displayed by the peptide system. The two mutant peptides **T9A** and **T26A** act as cross-catalysts for the formation of the native autocatalytic peptide but are autocatalytically inactive (taken from ref. 45).

More complex autocatalytic networks were built using these coiled-coil peptide structures,[13] including an example where a replicator and a more efficient analogue were made to compete for a common

starting material. The two templates showed a preference for mutualistic cross-catalysis over autocatalysis, which prevented the more efficient competitor from dominating the system.[47] An analysis of this system done by Ellington and co-workers[23] concluded that although cross-catalytic replicators could result interesting for an OoL scenario, mutualistic replicators where mutations do not result in discrete quantised changes (unlike it is the case for nucleic acid based replicators) could actually be incompatible with classic mechanisms of Darwinian evolution.

1.4.2. Peptide β-sheets

Self-replicating systems based on peptide β-sheets have also been reported. Exceptionally, product inhibition should not be a problem in these particular systems as the newly ligated peptide does not need to dissociate from the template to act as a catalyst. This feature may allow the system to approach exponential growth (provided the aggregates are disrupted regularly, e.g. by agitation). Ashkenasy and co-workers[48] have shown that amphiphilic peptide **1** comprising repetitive dyads of hydrophilic and hydrophobic amino acid residues (Glu-Phe) can form soluble, β-sheet aggregates in water. Furthermore, template-directed self-replication of **1** through NCL coupling of nucleophile **N** and electrophile thioester **E** was achieved (Figure 1.8).

1 **ABA-Glu-Phe-Glu-Phe-Ala-Cys-Glu-Phe-Glu-Phe-Glu-Pro-$CONH_2$**

N **ABA-Glu-Phe-Glu-Phe-Ala-$COSCH_2CH_2CONH_2$**

E **H_2N-Cys-Glu-Phe-Glu-Phe-Glu-Pro-$CONH_2$**

1^{gg} **ABA-Glu-Phe-Gly-Phe-Ala-Cys-Glu-Phe-Gly-Phe-Glu-Pro-$CONH_2$**

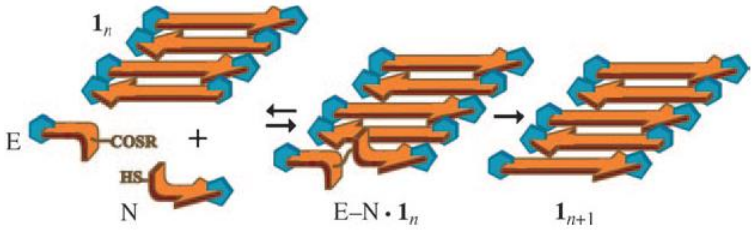


Figure 1.8: **Top**. Sequence of β-sheet forming peptide **1**, constituting fragments **E** and **N**, and a control template **1^{gg}**. **Bottom**. Schematic of the autocatalytic process of **1**: Antiparallel β-sheet **1_n** serves as a template for the association of **E** and **N**, which are coupled to form a new molecule of **1**. The larger aggregate **1_{n+1}** is then available for the next catalytic reaction (taken from ref 48).

A clear signature of this catalytic behaviour was observed while monitoring the ligation of the peptide fragments. Figure 1.9 shows that after a lag phase of about three hours, during which **1** is formed slowly, the rate of formation increases significantly; therefore, the assembly of **1** becomes more pronounced only after reaching a certain critical concentration. Moreover, reactions that were seeded with different concentrations of template **1** showed an enhancement of the initial ligation rates as the seeding

concentration was increased. This is the first self-replicating peptide reported that exhibits exponential growth.[49] A control reaction was performed by following the rate of formation of peptide **1^{gg}** (Figure 1.8), which hardly forms stable β-sheets in water owing to the introduction of two Gly residues in its sequence, and its production proceeded linearly. This system will be further analysed and discussed in chapter 2.

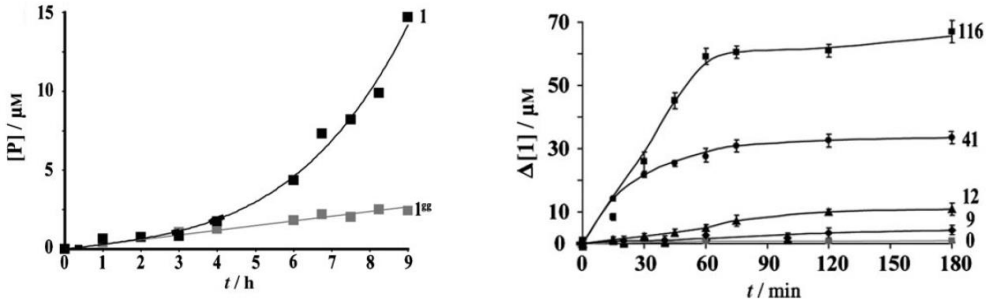


Figure 1.9: **Left**. Kinetic analysis of the ligation of peptides **E** and **N**, in black production of **1** over time, for comparison in grey the production of control peptide **1^{gg}**. **Right**. Production of **1** over time in reactions between **E** and **N** that were initially seeded with 9–116 μM amounts of **1** as a template (taken from ref 48).

In subsequent work, Ashkenasy[50] reported a system similar to self-replicator **1** where selection and amplification of certain members within a mixture of closely-similar peptides occurred. Peptide **2**, a close analogue of **1**, exhibited self-replication by promoting the ligation of an activated thioester (**E2**) and a free-NH_2 terminal fragment (**N2**). Due to the ligation mechanism (less controlled and selective than NCL) three isomers ("mutants") of **2** were generated as side products in this reaction: **2^{D}**, **2γ** and **2^{D}γ**, each bearing a modification (due to epimerization and/or non-native side chain coupling) at the Glu residue of the ligation site (Figure 1.10, top). Experiments where reactions between **E2** and **N2** were seeded with the different **2** isomers (individually) were carried out, and it was observed that in couplings seeded with canonical peptide **2** there was a threefold enhancement in the production rate of **2**, but there was little change in the production rate of the rest of the template isomers (Figure 1.10, bottom). Meanwhile, seeding experiments initiated with the three non-canonical **2** isomers induced an increase in the rate of production of canonical **2**, and lower or equal enhancement of the respective autocatalytic product.

$2_{isomers}$ **ABA-Glu-Phe-Glu-Phe-<u>Glu</u>-Phe-Glu-Phe-Glu-Phe-Glu-Pro-$CONH_2$**

N2 **ABA-Glu-Phe-Glu-Phe-<u>Glu</u>-$COSC_6H_4CH_2COOH$**

E2 **H_2N-Phe-Glu-Phe-Glu-Phe-Glu-Pro-$CONH_2$**

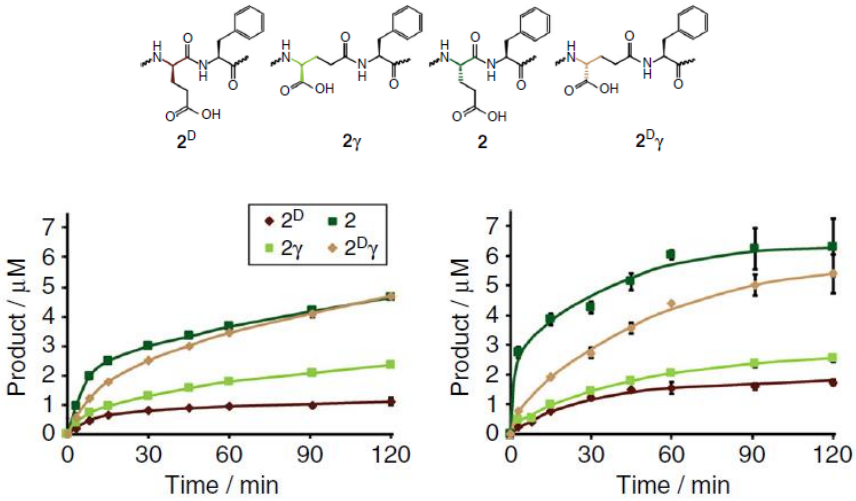


Figure 1.10: **Top**. Sequence of **2** and its fragments. The underlined Glu residue represents the location of the isomeric variation, the full structural representation of such variation is shown too. **Bottom**. Template assisted product formation in reactions between **E2** and **N2** seeded with different concentrations of the canonical template **2**: 30 µM (left) and 100 µM (right, taken from reference 50)

Given that this selectivity was a consequence of the non-canonical "mutants" acting more effectively as cross-catalysts rather than as autocatalysts, it was advocated as an "error correction mechanism" similar to that reported previously by Ghadiri (*vide supra*).

More recently, this type of replicators were upgraded from minimal systems by an example where a short peptide was capable of synthesising a complementary strand by templating the assembly of up to three of the complement's constituting monomers consecutively, in a stereo- and regioselective manner.[51] The system consisted of complementary peptide strands bearing binary alternating sequences of hydrophobic and hydrophilic amino acid residues (known to favour amphipathic β-strands). The individual template strand adopted random conformations in solution but once mixed with a complementary strand, the peptides assembled into insoluble β-sheet-like structures due to charge interactions. Template **$(FE)_4$** has a negative charge at neutral pH, while complementary substrate peptide **$R(FR)_3$** has a positive charge (and one Phe residue less than the template) which promotes their association (Figure 1.11, top). When activated Phe was added (stoichiometrically) to both the **$R(FR)_3$/$(FE)_4$** aggregate and to **$R(FR)_3$** alone, it was observed that the yield of the Phe addition product **$(FR)_4$** was much higher in the sample containing the co-aggregate, compared to the single peptide control (Figure 1.11, bottom). Double Phe additions of phenylalanine in the co-aggregate were not detected. Control reactions where different activated monomers were added to the system indicated the amyloid was sequence-specific for the addition of hydrophobic amino acids. Furthermore, when the

experiment was repeated adding activated DL-phenylalanine, the single addition showed a high preference for the isotactic product in the case of the co-aggregate, but that selectivity was lost for the single peptide **R(FR)$_3$**.

(FE)$_4$ **Ac-Phe-Glu-Phe-Glu-Phe-Glu-Phe-Glu-CONH$_2$**
(FR)$_4$ **H$_2$N-Phe-Arg-Phe-Arg-Phe-Arg-Phe-Arg-CONH$_2$**
R(FR)$_3$ **H$_2$N-Arg-Phe-Arg-Phe-Arg-Phe-Arg-CONH$_2$**
(FR)$_3$ **H$_2$N-Phe-Arg-Phe-Arg-Phe-Arg-CONH$_2$**
R(FR)$_2$ **H$_2$N-Arg-Phe-Arg-Phe-Arg-CONH$_2$**

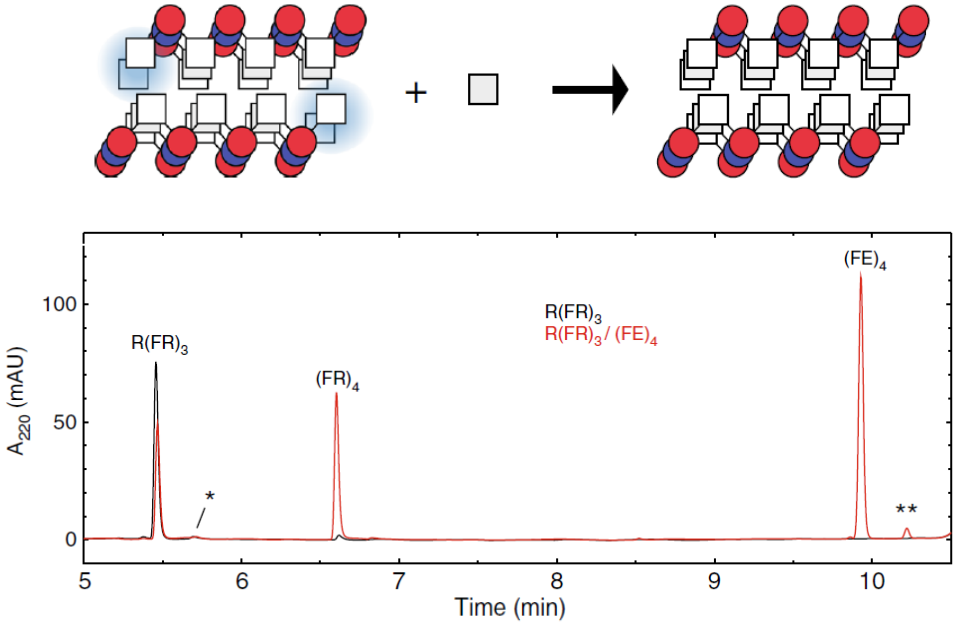


Figure 1.11: **Top**. Sequences of the complementary peptides. **Middle**. Schematic model of the amyloid-templated addition where Phe is represented by squares and Arg and Asp by blue and red circles respectively. Blue shading highlights the sites of addition in the substrate. **Bottom**. HPLC chromatogram of the Phe addition reactions. Reactions with just the substrate **R(FR)$_3$** are in black and those with the amyloid **R(FR)$_3$/(FE)$_4$** are in red (taken from reference 51).

To test the sequence specificity and consecutive templating capability of the system, stoichiometric amounts of different amino acid building blocks were added to a shorter substrate **(FR)$_3$** (two residues shorter than the complementary template) with and without the complementary template **(FE)$_4$** being present. It was observed that the amyloid enhanced the condensation of Arg while Asp was preferred by the substrate without the template. Moreover, when a mixture of activated Phe and Arg was added to **(FR)$_3$/(FE)$_4$**, the sequence-specific double addition product **(FR)$_4$** was observed (in low yield) but this product could not be detected in the absence of the template. A yet shorter substrate **R(FR)$_2$** was then treated with a mixture of activated DL-phenylalanine and DL-arginine (2 equivalents of each enantiomer) in the presence of **(FE)$_4$**, and it was observed that of 64 possible triple addition products, only three could be detected, the second mayor being **L-(FR)$_4$**. When the same mixture of racemic

amino acids was added to **R(FR)$_2$** without the **(FE)$_4$** template no **(FR)$_4$** could be detected. These results advocated for the sequence-selective templation capabilities of this amyloid.

Finally, the replicating peptides reported by Otto and co-workers[52] are worth mentioning briefly. These systems are unlike what has been discussed so far; they are based on a dynamic combinatorial library (DCL) of short peptides which form a variety of macrocyclic oligomers through reversible disulfide bonds. The macrocycles can then assemble into nanofibers held together by β-sheets. These nanostructures can self-replicate, competing for a common feedstock. The small DLC was made from self-binding building block **3** (Figure 1.12), which features a peptide sequence with alternating hydrophobic (Leu) and hydrophilic (Lys) amino acids. Peptide **3** formed a mixture of cyclic trimers and tetramers after stirring for four days. However, after that period, the composition changed and a cyclic heptamer became the dominant product, consuming both of the other macrocycles in the process. This behaviour was found to depend critically on the mode of agitation; the absence of mechanical agitation resulted in a mixture consisting of mostly the trimer and tetramer oligomers after long periods, but when shaking was used instead of stirring, the preferential formation of cyclic hexamer was observed. Circular dichroism (CD) and cryo-TEM analyses of solutions containing hexamer and heptamer macrocycles confirmed the presence of fibers originated from the peptide chains assembling into β-sheets (Figure 1.12). Notably, it was found that the formation of the larger macrocycles required a specific (β-sheet forming) peptide sequence.

The sigmoidal nature of the production of the hexamer and heptamer macrocycles suggested it proceeded autocatalytically. This was confirmed by experiments were small amounts of the hexamer and heptamer were seeded in samples of **3** which were shaken and stirred, respectively. In both cases, the formation of the corresponding seeded macrocycle was clearly induced in the samples, which consisted of mostly trimer and tetramer macrocycles at the point of addition.

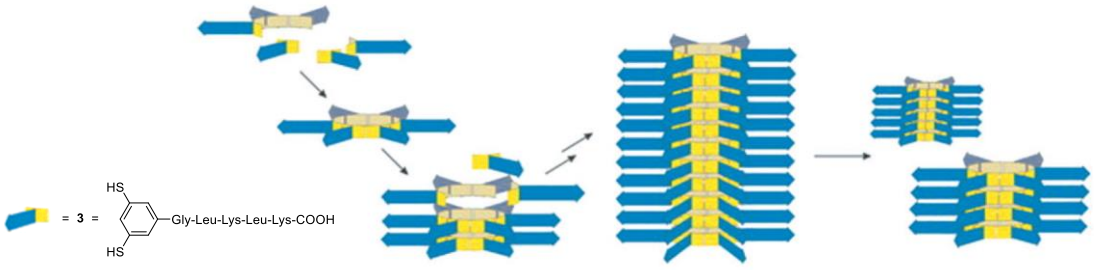


Figure 1.12: Structure of the monomer **3**, the benzenedithiol core shown in yellow and the peptide chain in blue. Schematic representation of the proposed formation of fibers of **3$_6$** (taken from ref. 52).

It was also found that the observed mechanosensitivity and kinetics were dependent on the fiber nanostructures. The exponential production of the macrocycles was explained by a templating effect occurring at the ends of the fibers aided by the mechanical breaking of such structures, which generated

more growing ends. Regarding the product selectivity, it was found that shaking selectively fragmented hexamer fibers while stirring the solution disrupted fibers composed of either hexamers or heptamers. Given that the heptamer growth was more efficient than that of the hexamer counterpart, the heptamers dominate the system when stirred. Thus, despite the system being under thermodynamic control, the product distribution was determined kinetically.

More recently, different versions of this system have been studied where changes to the sequence of the peptide building block were introduced (although it still consisted of residues with alternating polarities).[53] Networks of auto- and cross-catalytic replicators have also been reported in DCLs made from mixtures of different building blocks.[54]

In summary, peptide-based systems so far have demonstrated remarkable replicating capabilities and, in some cases, they have shown emerging properties emulating features found in biological systems. However, while some diversity is possible for a few of these replicators, the information storage capabilities of nucleic acids, originated from discrete residue-to-residue recognition, are still out of reach. Critically, plausible means of implementing a "digital-like"[19,55] coding fidelity similar to that displayed by RNA/DNA polymers, in (solely) peptide-based polymers are nontrivial.

1.5. Peptide Morse code

The notion that peptides could have co-evolved with or even preceded RNA before extant life is not new. The so-called "peptide/amyloid-world"[56,57] emerged as an OoL model due to the difficulties associated with RNA-first scenarios (*vide supra*) and considering the attractive properties of proteins regarding their building block's relatively easier prebiotic synthesis and their preponderant role in modern life. However, as it has been stated before, one of the main limitations for the emergence of selection and evolution in primitive peptide-based replicators comes from the inaccuracy of information transfer in standard peptides.

To overcome this problem, we propose the "peptide Morse code" (PMC): an informational system based on hybrid peptide β-sheets. It has already been demonstrated that β-sheet aggregates can act as templates for their own synthesis from shorter components. However, standard self-replicating peptide β-sheets are not suited for carrying information. Recognition in these systems is based on physicochemical interactions encoded in the amino acid side chain (normally a very specific sequence of units with alternating polarity). This makes the possibility of encoding a great deal of information in a residue-to-residue fashion difficult. Therefore, we have devised the potential to carry "digital-like" information, not through structural features encoded in the monomer side chain, but through hydrogen-bonding recognition in the peptide backbone. This is possible if both, α- and β-amino acid building

blocks are introduced in the polypeptide chain (Figure 1.13, top). In extended strands, the additional backbone methylene in the β-unit renders the two monomers incompatible, given that H-bonding sites are incorrectly aligned and they cannot optimally associate. This originates a discrete molecular recognition mechanism.

A simple variant of PMC built from glycine and β-alanine units is illustrated in Figure 1.13. This version is particularly attractive in a prebiotic context.[58] This figure shows the potential of α/β-peptides to form antiparallel β-sheets sequence-selectively, depending on the hydrogen-bonding complementarity of each amino acid unit. In the example, N to C sequence —•—••——• (• = glycine, — = β-alanine) will recognise and pair with complementary N to C sequence •——•••—•—, both strands alternating throughout the extended array. The use of short *vs* long units to encode information suggests the system to be described as "peptide Morse code" (PMC).

In principle, any sequence of building blocks is possible but naturally, there might be restrictions. Moreover, while some H-bonding between strands with mismatching α/β-sequences could still occur, we expect the reduced complementarity to make β-sheet formation much less favourable compared to matching sequences.

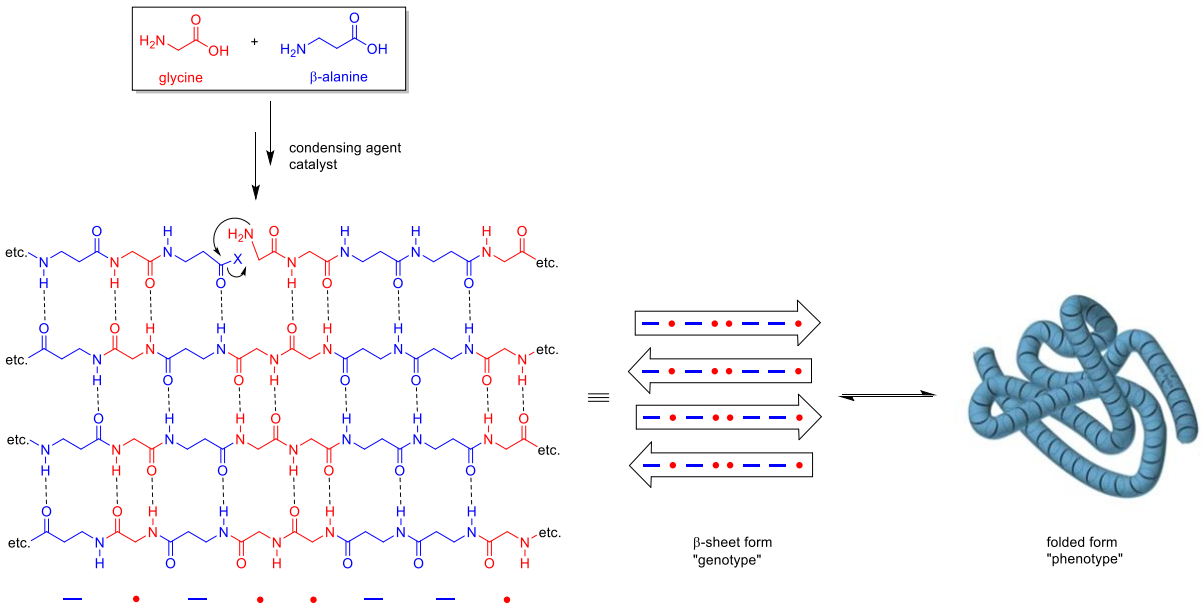


Figure 1.13: Prototypical informational self-replicating peptide β-sheet built from glycine and β-alanine.

The α/β-sequence-selective synthesis of new peptide strands from amino acid units or smaller peptide fragments can be templated by the β-sheet array (Figure 1.13, top), a process potentially mediated by catalysts or condensing agents. The aggregate could be in equilibrium with a folded version of the α/β-peptide strand, which may possess catalytic activity and could be considered as the phenotype resulting from the genotype encoded in the β-sheet (Figure 1.13, right). It could be postulated that such enzymatic behaviour might include the ability to promote its own synthesis. Alternatively, it could be the β-

aggregate itself possessing the catalytic activity (phenotype) depending on the morphology of higher cross-β assemblies.[19]

These are some of the most attractive features of β-sheet PMC as an informational self-replicator:

- Amino acid building blocks are considered to be the biomonomers easiest to achieve in experiments recreating prebiotic earth conditions (this is also true for some non-proteinogenic β-residues),[7,58–61] they have also been found in meteorites[62] and feasible pathways for their oligomerization into peptides under such conditions have also been described.[57,63–66]

- The β-sheet aggregates can serve as templates for replication without the need for strand separation (see section 1.4.2), overcoming the product inhibition problem inherent to closed duplex-forming systems (such as nucleic acids) therefore enabling amplification.

- α/β-Peptides that assemble into β-sheet-like structures in aqueous (and organic) media are known (*vide infra*)., although, using the α/β-amino acid sequence to encode information has not yet been explored. In this respect, it is also of note that both glycine and β-alanine oligomers have been found to assemble into extended sheet-like structures in the solid state.[67,68]

At this early stage, several elements of our hypothesis remain unclear. The medium in which self-replication would take place is still undefined. Aqueous media would be the most obvious choice, but other media and phases such as liquid-solid surfaces might have also been available for this process to take place. This would also be a determinant factor in the nature of the amino acids involved and the (activation) oligomerization process.

Furthermore, the prototype version shown in Figure 1.13 is not the only one to be considered. The system allows for variation so that a range of substituted amino acids could participate (Figure 1.14). In principle, the binary system based on two amino acids (glycine and β-alanine) can be fault-tolerant, which means that if an unwanted component is introduced (to some extent), it may not prevent the accurate transfer of information. However, this would alter the folded structure, disrupting the phenotype and tampering with the selection process. On the other hand, provided that a mechanism exists to ensure that just one amino acid of each type (α and β) predominates,[6,65] a number of combinations of binary systems would be possible.

Other possible boundaries of the system are hard to visualise *a priori*; specifically, those related to α/β-sequence tolerance, the fidelity of the replication process and the maximum length of the polypeptide allowed, as well as the correlation of this length to the replication efficiency and fidelity.

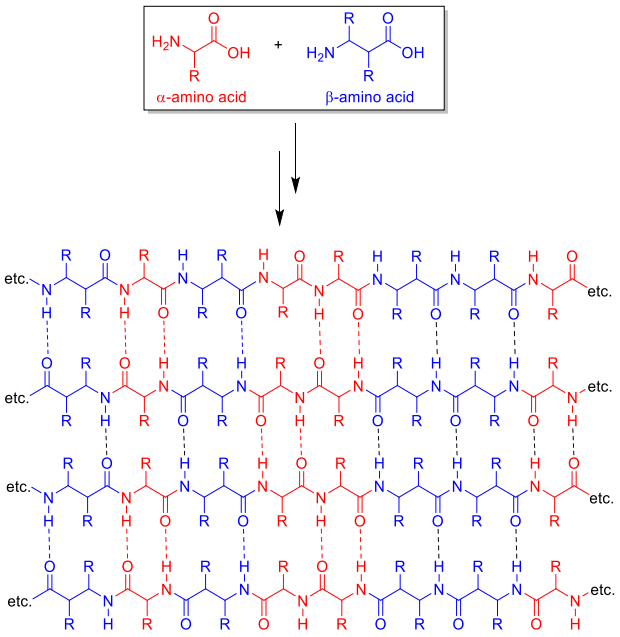


Figure 1.14: A more general representation of a PMC β-sheet made from undefined sets of α and β amino acids.

Finally, it is worth noting that our hypothesis does not attempt to disprove the RNA-world but merely study the plausibility of informational self-replication on a persuasive alternative system made from relatively abundant building blocks in prebiotic Earth and, further down the line, study the plausibility of the emergence of selection and Darwinian evolution.

1.6. β-Sheet-like assemblies from α/β-peptide hybrids

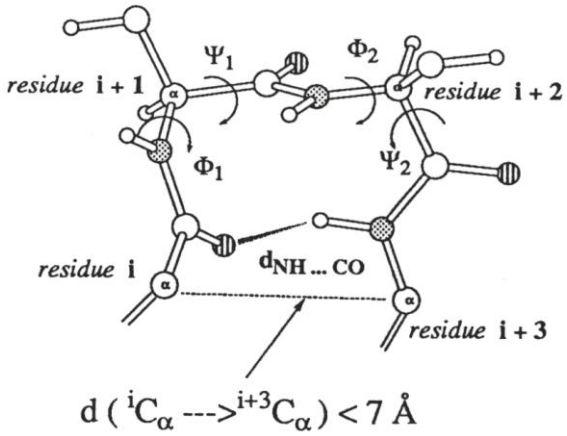
The synthesis and structural characterization of β-peptides[ii] has received considerable attention over the past decades; mainly with the aim of exploring the novel folded structures and conformation modes displayed by these polymers and finding an application for them in *de novo* protein design.[68] Moreover, given that the incorporation of β-residues into α-peptide structures has been shown to confer interesting biological activity to the mixed (or hybrid) peptides, the study of these α/β-hybrid sequences has also attracted attention.[69] Several protein-mimetic structures from these hybrids have been reported, including helices, turns and extended sheet-like conformations.

[ii] Peptides composed of β-amino acids

1.6.1. α/β-Peptide hairpin structures in organic media

Given the importance of this type of secondary structures to the present work (as it will be seen in chapter 3), a brief description of the peptide β-hairpin motif is in order.

β-Hairpins are among the most important regular secondary structure motifs found in nature.[70] They are composed of two consecutive hydrogen-bonded antiparallel β-strands connected by a turn or loop sequence. Turns are segments of a polypeptide where the peptide chain reverses its overall direction. They are typically composed of four (β-turn) amino acid residues. β-Turns are classified according to the values of four backbone torsional angles (φ_1, ψ_1, φ_2, ψ_2, Figure 1.15). Pro is the most common amino acid in type I and type II turns because of the restriction of φ to about -60°.[71]



Turn type	φ_1,	ψ_1	φ_2	ψ_2
I	-60	-30	-90	0
II	-60	120	80	0
I'	60	30	90	0
II'	60	-120	-80	0

Figure 1.15: Ideal backbone torsion angles of the common β-turn types (taken from reference 71).

Some of the earliest reports of α/β- peptide hybrids adopting β-hairpin conformations involve the use of β-residues such as β-alanine and dinipecotic acid as turn-nucleating motifs bearing all-α strands.[72,73] Other examples involved hairpins with short β-strands composed of β-residues, nucleated by an all-α turn.[74] However, the most relevant structures to our work are those incorporating mixed α/β-sequences in the antiparallel sheet-like β-strand segments.

The earliest reports of such peptides were done by Balaram and co-workers.[72,75] They synthesised a pair of short α/β-peptides bearing a ^{D}Pro-Gly β-turn motif (Figure 1.16, **4** and **5**), which were shown to assemble into β-hairpins in organic solvents. Peptide **4**, containing β^3hVal and β^3hLeu residues at the *i* and *i* + 3 positions of the turn (see Figure 1.15), was analysed by ROESY NMR spectroscopy in CD_3OD, the observed sequential $d_{\alpha N}$ ($C^{\alpha}H\leftrightarrow NH$ NOE contact) and a cross-strand d_{NN} NOE support an extended conformation for the β-strands (Figure 3a, bottom). The NOE pattern observed at the β-turn suggested a type-I' structure. Similarly, decamer (10-mer) **5**, bearing two β^3hPhe residues opposite to each other in the β-strands, was analysed by ROESY NMR spectroscopy in CD_3OD displaying four cross-strand NOEs distributed along the strands, providing strong support for the peptide adopting a β-hairpin

conformation. The peptide hybrids were also analysed by CD in methanol and their spectra displayed a strong negative band between 218 and 220 nm, characteristic for the β-conformation. Furthermore, single crystal X-ray diffraction analysis could be carried out for both peptides, and the data supported the formation of type-I' nucleated β-hairpin structures. Notably, extended antiparallel β-sheet arrays were observed in the crystal structure.

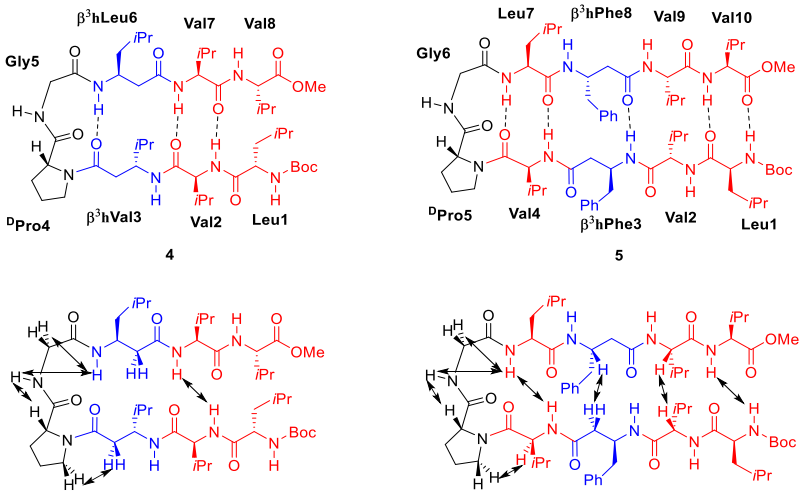


Figure 1.16: **Top**. Structure of the short α/β-peptide hybrids **4** and **5** adopting a β-hairpin conformation. **Bottom**. Schematic representation of relevant NOEs indicated by double edged arrows (left **4**, right **5**).

In a subsequent report,[76] a new pair of α/β-hybrids were studied. This time both hybrid 8-mers; **6**, a shorter version of **5** bearing a pair of β^3hPhe[iii] residues and **7**, bearing four β-residues consisting of pairs of β^3hVal and β^3hLeu residues (Figure 1.17). The aim of these designs was to study the effect of different orientations in opposing residues bearing benzyl side chains. In a similar fashion to the previous pair of peptide hybrids, NMR and CD studies were carried out to characterise the conformation of these molecules in solution. In the case of peptide **6**, ROESY NMR analysis in CD_3OH showed large $d_{\alpha N}$ contacts (for α-residues), consistent with extended strand conformations, a single cross-strand contact near the β-turn (a type II') also supported the folded structure. However, the absence of cross-strand NOEs expected further away from the β-turn and the presence of Leu1 NH↔Val8 $C^{\alpha}H$ and Leu1↔β^3hPhe2 (d_{NN}) indicated that local helical conformations may also be populated in solution. In the case of **7**, the observation of three cross-strand NOEs supports an antiparallel hairpin structure. The pattern of contacts observed at the turn region suggests both type I' and II' conformations are populated. However, a series of weak d_{NN} were also observed for this peptide, which were incompatible with a completely rigid hairpin structure, suggesting some fraying of the strands near the N- and C-terminus. The circular dichroism spectrum of **6** showed a broad negative band with a minimum at 224 nm. On the other hand, peptide **7** produced an anomalous CD spectrum with a negative band at 234 nm and a

[iii] The superscript "3" indicates that the side chain extends from the backbone carbon adjacent to the nitrogen. The "h" stands for homo.

positive band at 221 nm. The origin of this red shift was rationalised as an interaction between the phenyl chromophores of the Phe residues.

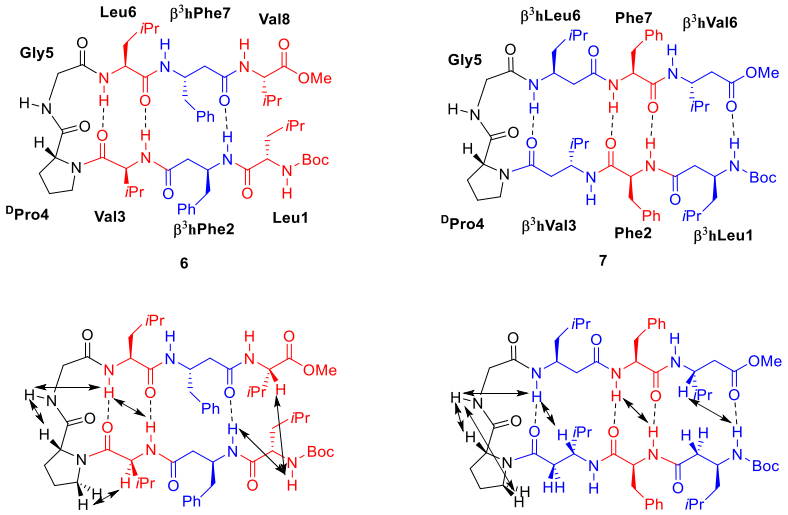


Figure 1.17: **Top**. Structure of 8-mer α/β-peptide hybrids **6** and **7** adopting a β-hairpin conformation. **Bottom**. Schematic representation of relevant NOEs indicated by double edged arrows (left **6**, right **7**).

In a different study by the group, an analogue of 10-mer **5** where the β^{3}hPhe residues were replaced by β-alanine residues was synthesised. ROESY NMR spectroscopy and CD both supported a predominant population of a β-hairpin conformation in solution (CD_3OH), although possibly with some degree of fraying near the termini.

Finally, Balaram[77] has also shown that these folded structures made from antiparallel α/β-peptide strands can be expanded by introducing two β-turn-nucleating ^{D}Pro-Gly segments, generating multi-stranded β-sheet peptide. 19-mer **8** was designed taking 10-mer **5** as a base, adding a second β-turn and elongating the strands (Figure 1.18, left). Peptide **8** contains three β^{3}hPhe residues, aligned in the antiparallel strands. NMR analysis in methanol allowed the calculation of a structural conformation for this peptide. Several techniques including temperature dependence of NH chemical shifts, rates of H/D exchange in amide protons, comparison of vicinal constants and critically cross-strand NOE interactions (Figure 1.18, right) were employed to characterise the structure adopted by **8**.

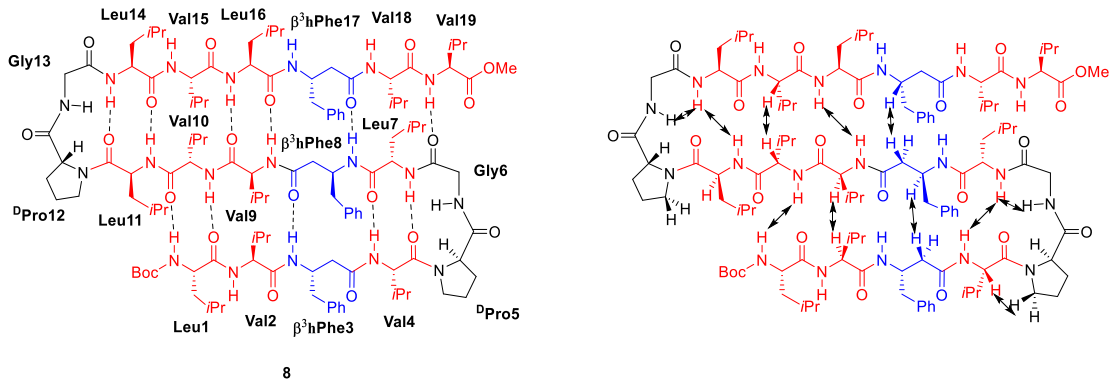


Figure 1.18: **Left**. Structure of 19-mer α/β-peptide hybrid **8** adopting a multi-stranded β-hairpin conformation. **Right**. Schematic representation of relevant NOEs indicated by double edged arrows.

1.6.2. Extended pleated-sheet assemblies of amphiphilic α/β peptides

Extended β-plated sheet assemblies made from α/β-peptides have precedent in the literature too. Gellman and co-workers[78] have reported the design of a pair of amphiphilic 11-mer α/β-peptides bearing a 1:1 alternating pattern of α and β subunits along the backbone (Figure 1.19).

The peptide sequences are defined by an alternating pattern of hydrophobic dyads (β^3hPhe-αVal) and hydrophilic dyads (β^3hGlu-αThr or β^3hLys-αThr). This sequence pattern causes the hydrophobic side chains to project from one side of the extended backbone and the hydrophilic side chains to project from the other side, thus giving rise to global amphiphilicity. This was expected to promote sheet assembly at the air–water interface, as the hydrophilic side chains can project into the aqueous phase while the hydrophobic side chains project into the air. Notably, the sequential arrangement of hydrophobic and hydrophilic side chains giving rise to amphiphilicity in an extended conformation is different for these α/β-peptides relative to conventional α-peptides; in the latter case, hydrophobic/hydrophilic side chains are arranged in a 1:1 alternating pattern. A small capping group was placed at the N-terminus and the C-terminus was functionalised as an amide to prevent development of charge at these positions. The terminal Pro residues are intended to promote the alignment of α/β-peptide strands along the H-bond direction. All the α-residues (Val and Thr) have a branched side chain. This design feature is based on previous findings that such residues strongly destabilise helical secondary structure among α/β-peptides with a 1:1 α/β pattern (Figure 1.19).[79]

The two variants examined, **βEβK** and **βKβE** differ in the positions of the ionizable side chains. Grazing incidence X-ray diffraction (GIXD) measurements at the air–water interface showed difference in the patterns presented by both sequences, which indicated that peptide **βEβK** adopted a more ordered β-sheet structure than peptide **βKβE**. It was hypothesised that this difference arises from the difference in the interaction of the charged side chains with the molecular dipole of the α/β-peptide backbone in the

pleated strand conformation. A model of the crystalline assembly formed by **βEβK** based on the data obtained from the diffraction data (and some basic conformational assumptions) was generated (Figure 1.19).

βEβK peptide **Ac-Pro-β³hPhe-Val-β³hGlu-Thr-β³hPhe-Val-β³hLys-Thr-β³hPhe-Pro-NH₂**

βKβE peptide **Ac-Pro-β³hPhe-Val-β³hLys-Thr-β³hPhe-Val-β³hGlu-Thr-β³hPhe-Pro-NH₂**

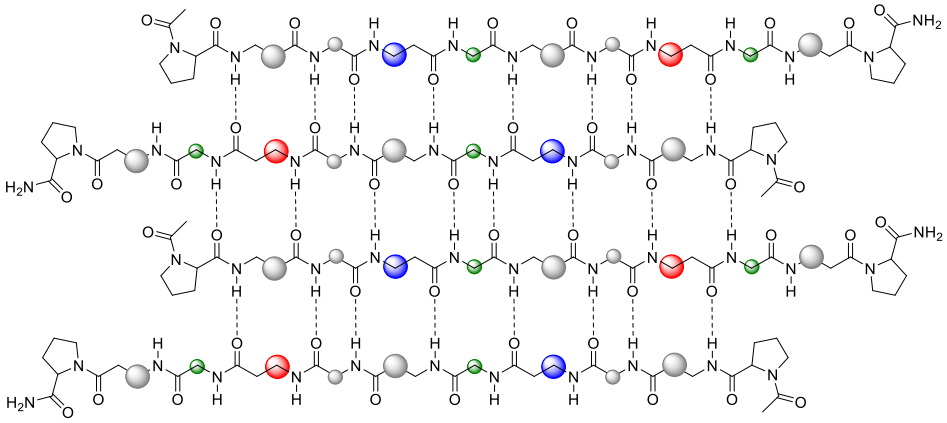


Figure 1.19: **Top**. Sequence of the two 11-mer α/β-peptides **βEβK** and **βKβE**. **Bottom**. Proposed two-dimensional lattice of the amphiphilic oligomer **βEβK** packed in antiparallel mode. Circles denote side chains of amino acids, hydrophobic amino acids, β³hPhe (large gray circles) and Val (small gray circles), hydrophilic amino acids β³hLys (blue), Thr (green), and β³hGlu (red) circles, respectively).

Subsequently, this same research group reported another pair of amphiphilic 11-mer α/β-peptides analogues **βE₂** and **βK₂**.[80] These new peptides preserved the same sequence pattern of the two predecessors and differed only in the identity of the charged residues; **βE₂** a dianion bearing two β³hGlu residues and **βK₂** a dication bearing two β³hLys residues. GIXD and FT-IR measurements indicated that peptide **βE₂** formed ordered assemblies similar to those formed by **βEβK**, which was not the case for peptide **βK₂**. This difference was attributed to the β³hGlu side chain carboxyl groups in **βE₂** not being ionised and possibly engaging in interstrand hydrogen-bonding, which contributed to the ordered strand assembly.

1.7. Scope of thesis

In this introduction, the OoL problem and some models to tackle it have been overviewed. In that context, we have stated the relevance of a peptide-based informational self-replicator and introduced the PMC hypothesis. Several literature peptide replicators have been discussed and β-sheet-like structures formed by α/β-peptide hybrids have been analysed to provide a background to our model.

This thesis focuses on achieving a proof of concept for PMC. To our knowledge, a solely peptide-based polymer designed to encode information through hydrogen-bonding patterns in the peptide backbone is unprecedented. Therefore, these first studies, aimed at gaining proof of the feasibility of this approach, have used a number of design features simplifying our system and biasing it towards β-sheet formation and template-directed ligation. Such features include the use of palindrome sequences, the use of β-sheet breaker residues or turn segments, the use of highly chemoselective ligation reactions or ligation reactions carried in non-nucleophilic solvents and the monitoring of peptide conformation in organic media.

Chapters 2 and 4 are dedicated to investigating sequence-selective self-replication of α/β-peptide hybrids. The first design is based on a literature water-soluble replicator and the second one adapted from successful results obtained working in organic media throughout this project.

In chapter 3, we focus on obtaining structural evidence for the sequence-selective formation of β-sheet-like structures by α/β-peptide hybrids. To achieve this, we used a simplified intramolecular model based on the β-hairpin motif, monitoring conformational differences displayed by molecules containing matching *vs* mismatching α/β-sequences in organic media.

Finally, in chapter 5 we explore the intramolecular templating effect arising from α/β-sequence recognition in the antiparallel β-strands of a β-hairpin.

Chapter 2. Design of a water-soluble self-replicating PMC system

Unless stated otherwise, all results were obtained by the author of this thesis (Alberto Avila Castro)

2.1. Background

2.1.1. A self-replicating amphiphilic β-sheet peptide

As mentioned in the previous chapter, Ashkenasy and co-workers[48] have shown that amphiphilic 12-mer peptide **1** can form soluble β-sheets using CD, DLS and TEM. Furthermore, **1** was shown to accelerate its own synthesis from the NCL coupling of nucleophile **N** and electrophile thioester **E** (Scheme 2.1). Signature reaction rate enhancement was observed when the ligation of the fragments was carried out in the presence of different concentrations of template **1** seeds.

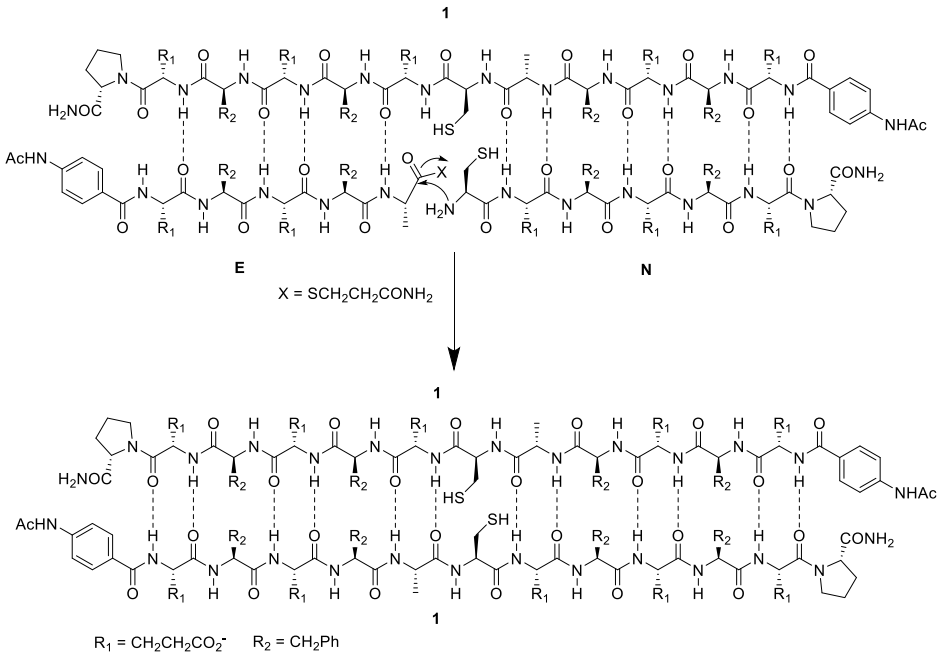


Figure 2.1: Structure of β-sheet forming peptide **1**, and self-replication reaction through NCL of peptide fragments **E** and **N** in aqueous solution.

Further studies analysed the self-assembly process experimented by peptide **1** and correlated the transition between various supramolecular structures to the observed autocatalytic activity. It was found that the initial β-sheet structures assemble into larger supramolecular nanostructures including fibers and nanotubes (Figure 2.2). The fibers ($\mathbf{1_F}$), which are believed to be the catalytically active species, were shown to exist only transiently during this process. On the other hand, the nanotube structures formed at later stages were found to be inactive.

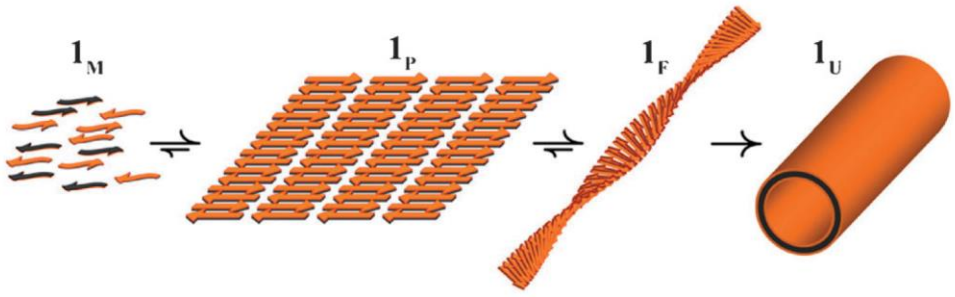
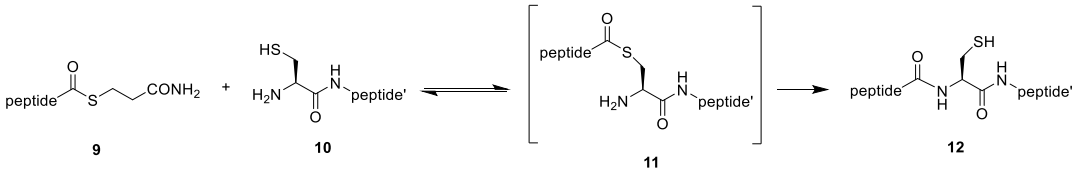


Figure 2.2: Supramolecular mechanism of β-sheet replication. The sheets (**1P**) assemble into fibers (**1F**) and then nanotubes (**1U**). The transient fibers are believed to be the active species (taken from ref 49).

2.1.2. Native chemical ligation

Taking into consideration its use in the present study, a brief overview of NCL should be given. This process represents the most widely used chemoselective peptide ligation strategy. The chemoselective capture step is mediated by a reversible thiol-thioester exchange between an electrophilic thioester at the C-terminus of peptide fragment **9** and the nucleophilic thiol of a Cys residue located at the N-terminus of peptide fragment **10** (Scheme 2.1). In the following rearrangement, the Cys-thioester **11** undergoes a rapid S to N shift, *via* a favourable five-membered transition state, to form a native peptide bond between the two peptide fragments.[81]

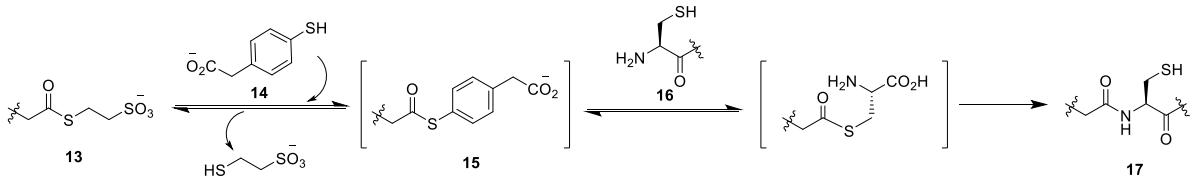


Scheme 2.1: General representation of the native chemical ligation (NCL) strategy.

A major advantage of NCL is the mild reaction conditions, which allow the ligation to be carried out in buffered aqueous solutions of neutral pH value. This is important for two reasons: on the one hand, thioesters are not stable under basic conditions. On the other, acidic conditions do not favour NCL because the reactivity of the N-terminal amine of the nucleophilic peptide fragment is reduced. Another important factor during the NCL reaction is the nature of the C-terminal amino acid residue of the electrophilic fragment. The fastest reaction rate was observed for the Gly thioester which reacted quantitatively in less than four hours. In contrast, thioesters containing β-branched amino acids or proline did not result in a quantitative conversion, even after two days.

Finally, the nature of the thioester has a huge impact on the NCL reaction. Thioalkyl esters **13** (Scheme 2.2) are commonly used due to their ease of preparation and handling, although they react relatively

slowly under NCL. This problem is normally circumvented by performing ligations in the presence of an aromatic thiol additive **14** to promote *in situ* transthioesterification to a more active thioester intermediate **15** which then can ligate faster to the Cys-terminal peptide fragment **16** and generate product **17**.[82]



Scheme 2.2: *In situ* thiol–thioester exchange during NCL.

2.2. Project aims

To investigate the ability of Morse code peptides to replicate sequence-selectively through the formation of covalent bonds, we envisioned to use the Ashkenasy replicating peptide as a model (see section 1.4.2). Replicator **1** (see Figure 2.1) is not informational, and it is not obvious how information could be encoded using only α-amino acid components. However, it could be used to demonstrate the PMC concept by building an α/β-peptide hybrid analogue (Figure 2.3). Fragments of this hybrid template bearing complementary sequences of α/β-residues and appropriate reactive termini should be recognised by it and ligated in an auto-catalytic fashion (Figure 2.3). Seeding the ligation of such fragments with the α/β-template should have an accelerating effect on the ligation rate. In further experiments, other variants of this α/β-template bearing a different sequence of α/β-residues would be synthesised. Then, it could be shown that seeding alternative template variants with non-complementary fragments would result in a lack of recognition (or at least reduced recognition) and therefore no templating effect.

This would serve as a demonstration of the potential for peptide Morse code to transmit information through the hydrogen-bonding complementarity encoded in the sequence of α/β-residues. If these first experiments were successful, more intricated and more prebiotically-interesting variants would be designed and studied.

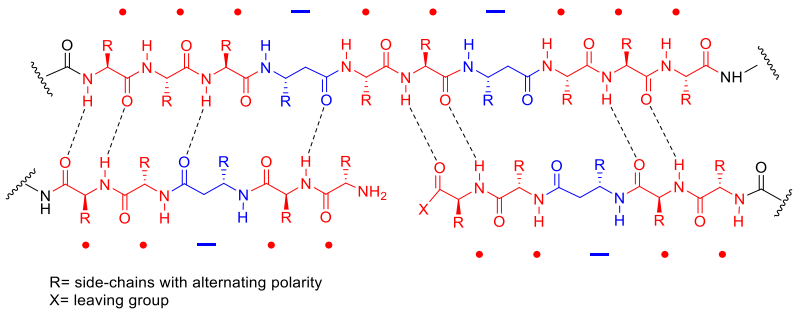


Figure 2.3: α/β-variant of a β-sheet peptide replicator. A template (**top**) recognizing fragments with complementary α/β-sequences (**bottom**).

2.3. Amphiphilic PMC system

2.3.1. System design

As previously mentioned, Ashkenasy's α-peptide **1** was designed to form β-sheet aggregates in water, and was shown to catalyse its own synthesis from a pair of corresponding fragments *via* NCL.[48] Figure 2.4 illustrates an α/β-analogue of **1** (**18**). The molecule preserves the facial amphiphilicity and the central Ala-Cys linkage but incorporates two β-amino acids. These β-homoglutamic acid and β-homophenylalanine residues are positioned so that peptide **18** is self-complementary, i.e. palindromic, and thus able to form antiparallel β-sheets alternating with itself. The terminal proline residue was used since it does not possess a backbone N-H group available for hydrogen-bonding and hence it can act as a β-pleated sheet breaker, promoting the formation of ordered assemblies. Capping of the N-terminus with the aromatic 4-acetamidobenzoic acid (ABA) label was carried out to facilitate high-performance liquid chromatography (HPLC) monitoring of the ligation reaction (peptides containing ABA are detected at 270 nm).[48]

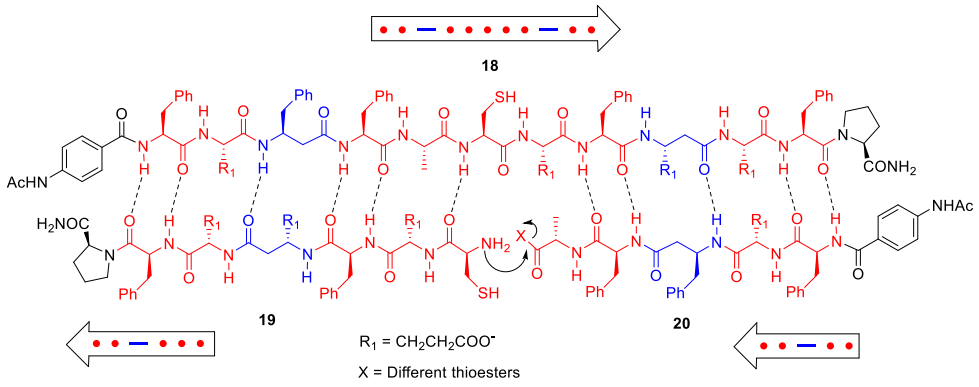


Figure 2.4: Design of the self-replicating amphiphilic PMC system.

Template **18** should be generated from the coupling of aminothiol **19** and thioester **20** fragments by NCL, potentially in a template-directed fashion (Figure 2.4). Ligation of fragments **19** and **20** should be accelerated in the presence of template **18**, but addition of a peptide template with a non-complementary sequence of α/β residues (e.g., **1**) should have little or no effect.

2.3.2. Peptide synthesis

The sequences of template **18**, nucleophilic fragment **19** and C-terminal carboxylic acid peptides **21a-b** (later to be converted into activated thioesters) are illustrated in Figure 2.5. The replacement of the Ala residue in the original design for a Gly residue in fragment **20b** will be discussed later (*vide infra*).

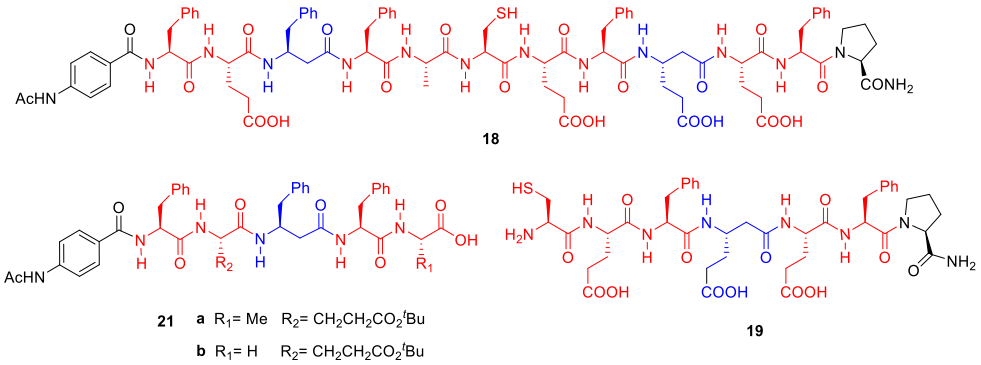
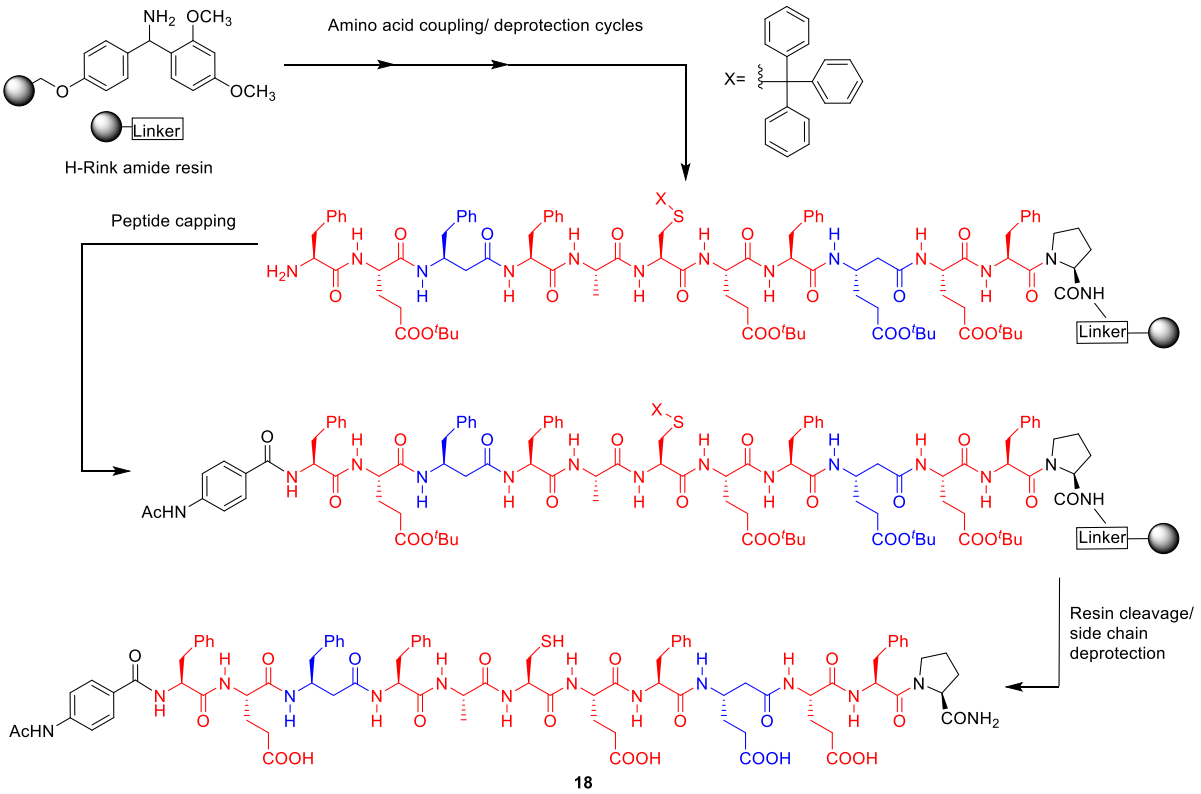
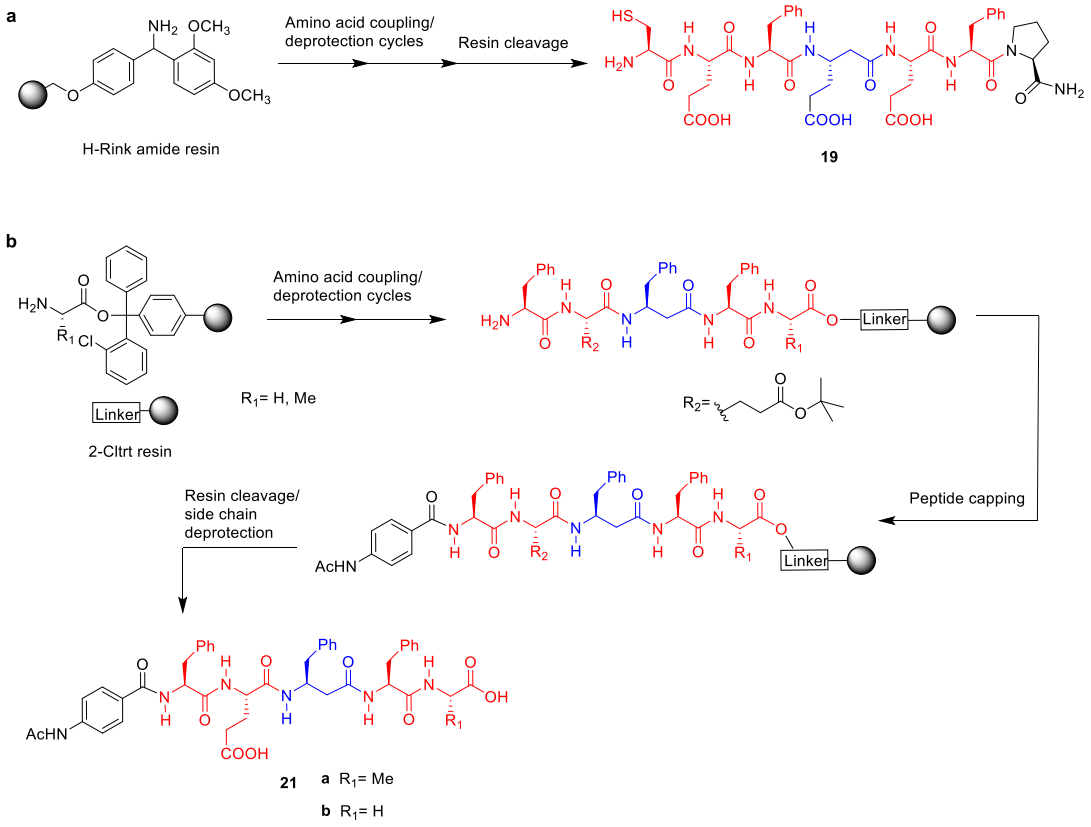


Figure 2.5: Structure of peptide template **18** and fragments **19**, **21a** and **21b**.

These molecules were synthesised using a standard Fmoc-based manual solid-phase peptide synthesis (SPPS) methodology (see also section 7.2.1). An H-Rink amide resin was used to obtain peptides **18** (Scheme 2.3) and **19** (Scheme 2.4), while H-Ala-2-chlorotrityl and H-Gly-2-chlorotrityl resins were used to synthesise the C-terminal free carboxylic acid fragments **21a** and **21b**, respectively (Scheme 2.4). The peptides were purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)). The identity and purity of the peptides was assessed by MALDI-TOF MS and analytical HPLC (in the case of **21a**, also by ^{1}H and ^{13}C NMR).



Scheme 2.3: Overview of the solid-phase peptide synthesis (see section 7.2.1) of peptide **18**.

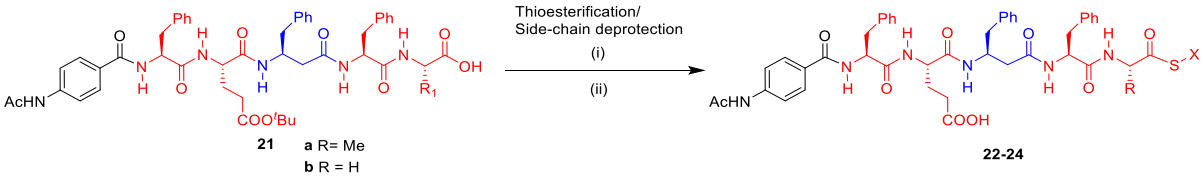


Scheme 2.4: Overview of the solid-phase peptide synthesis (see section 7.2.1) of **19** (**a**), **21a** and **21b** (**b**).

2.3.3. Synthesis of peptide thioesters

Once peptide fragments **21a-b** were obtained, the C-terminal carboxylic acids were converted into thioesters suitable to use in NCL experiments (**22-24**, Tables 2.1 and 2.2). A range of commonly used thioester derivatives were synthesised. We rationalised that this would allow us to tune the NCL reaction in order to demonstrate the self-replicating effect. Thioesterification was achieved by treating the peptides with the desired thiol using PyBop as activating agent and DIPEA as base.[83] The reactions were monitored by analytical HPLC and stopped once the peptide precursor had disappeared. After thioesterification, side chain deprotection of the Glu residue was performed. The peptide thioesters were isolated in moderate to good yields after purification (Table 2.1). Poor yields of **23** and **24** are mainly attributed to low recoveries from HPLC purification. Notably, the thiophenol derivative could not be identified by MALDI-TOF MS (Table 2.1, entry 4), instead a compound with a 910.9 mass-to-charge ratio was detected, corresponding to the loss of water from the target molecule. Additionally, this compound showed very poor solubility, therefore it was not used.

Table 2.1: Synthesis of thioester derivatives **22-24**.



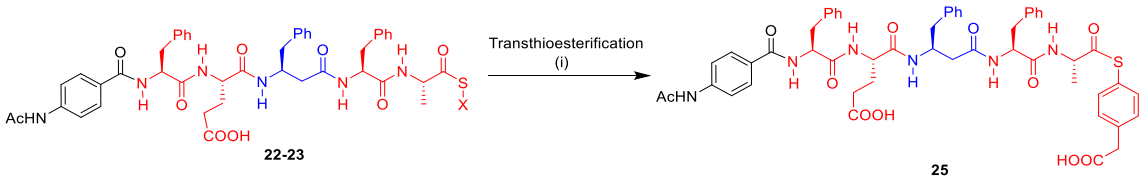
Entry	R_1	X	Compound	Yield[a]
1	Me	Bn	**22**	62%
2	Me	$CH_2CH_2SO_3Na$	**23**	38%
3	H	$CH_2CH_2SO_3Na$	**24**	36%
4	Me	Ph	-	Not observed[b]

(i) R_1SH, PyBop, DIPEA, DMF, 1.5 h. (ii) TIPS/H_2O/TFA (2.5:2.5:95), 2 h. See section 7.3.1
[a] Isolated yields from semi-preparative HPLC.
[b] Molecular ion was not observed by MALDI-TOF MS.

We were also particularly interested in trying thioester peptide **25** (Table 2.2) in NCL, considering that 4-mercaptophenylacetic acid (MPAA) has been described as a highly effective catalyst in this reaction.[84] To avoid a potential disruption of the catalytic β-sheet templation under study caused by generating the active thioester *in situ* (see section 2.1.2), it was decided to pre-form the active thioester derivative **25** instead.[85] Given that MPAA features a free carboxylic acid in its structure, the standard methodology used for the synthesis of thioester derivatives **22-24** could result in undesirable self-coupling reactions. Instead, thioester **25** was obtained through transthioesterification of **22** and **23** with

an excess of MPAA (Table 2.2).[85] The benzyl mercaptan precursor **22** afforded the best yields of thioester **25** (Table 2.2, entry 2).

Table 2.2: Synthesis of thioester derivative **25** from thioesters **22** and **23**.



Entry	X	Yield of 25[a]
1	Bn	85%
2	$CH_2CH_2SO_3Na$	20%

(i) MPAA excess, 6 M GdmCl/ 0.2 M phosphate buffer at pH = 7.1. See section 7.3.1
[a] Isolated yields from semi-preparative HPLC.

Before studying the coupling of our peptide fragments by analytical HPLC, a qualitative evaluation of the electrophiles **22**, **23** and **25** was carried out by MALDI-TOF MS. In all cases, upon mixing the electrophiles with peptide fragment **19** at pH 7 a signal at 1752.2 *m/z* corresponding to product **18** $(M+Na)^+$ was detected in the reaction mixture. With these results and knowing that MPAA forms very active thioesters,[50] it was decided to assess NCL of electrophilic fragment **25** by analytical HPLC.

2.3.4. Native chemical ligation using electrophile 25

NCL of fragments **19** and **25** (Figure 2.6 and Table 2.3) was monitored over 48 hours. Figure 2.6 shows representative HPLC traces of the reaction mixture at three different sampling points (analytical conditions are described in section 7.4.1). It is worth to point out that nucleophilic fragment **19** is not observed on the 270 nm channel as the fragment does not contain the ABA moiety. From the beginning of the ligation, it was observed that the MPAA derivative **25** disappeared from the chromatogram (t_R = 7.8 min) and instead a compound eluting very early (2.8 min) appeared in the 270 nm channel. As the reaction proceeded and reductive additive TCEP·HCl (t_R = 2.6 min) was consumed, nucleophile **19** slowly oxidised forming a disulfide-bonded dimer.[iv] Furthermore, a very low intensity peak corresponding to the ligation product **18** (t_R = 12.1 min) was observed in the HPLC traces throughout

[iv] This disulfide dimer of **19** was generated separately by stirring **19** in buffer (pH = 7) for 48 hours in an open flask. The dimer was then analysed by HPLC (t_R = 6.9 min) and MALDI-TOF (observed $[M+Na]^+$ = 1845.4 Da).

the reaction (<1%). Product **18** could also be detected in the reaction mixture by MALDI-TOF MS. These observations pointed to the ligation occurring only at very low conversions.

Attempting to improve the conversion of the ligation, different conditions were screened. The reaction temperature, final pH, peptide concentration and reductive additive loading were varied (Table 2.3). However, product **18** was still barely detectable in the chromatogram (Table 2.3, entries 1-7). These results along with a control experiment where fragment **25** was treated with TCEP·HCl in the absence of nucleophile **19** led us to conclude that **25** was being degraded by the phosphine additive, and this was tampering with the ligation. Unfortunately, the reaction would not proceed in the absence of TCEP·HCl, even when using degassed solvent (Table 2.3, entries 8-12). Alternatively, the reaction was carried out in the presence of an excess of MPAA (an excess of small thiols has been reported to prevent Cys dimerization in NCL[86]), but no significant improvement was achieved (Table 2.3, entries 13 and 14).

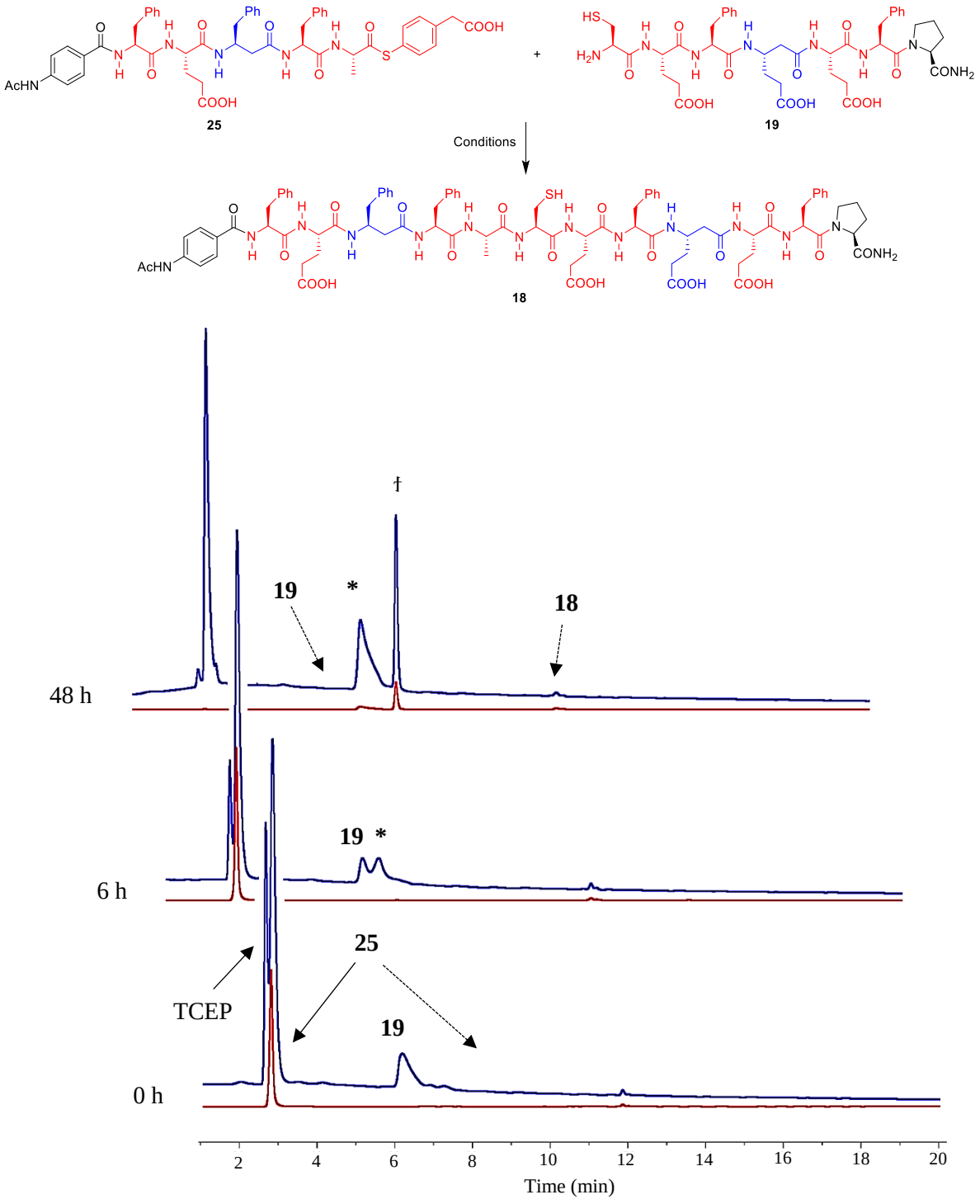
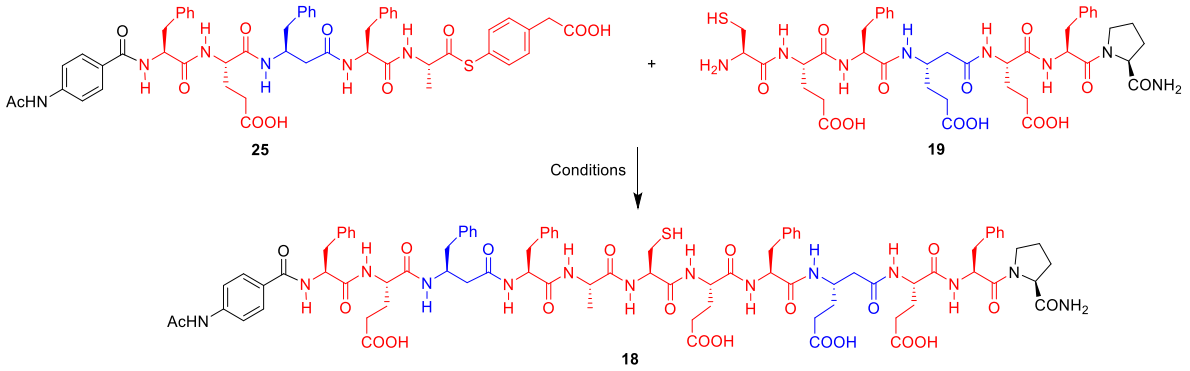


Figure 2.6: **Top**. Scheme of the NCL of peptides **19** and **25** (reaction conditions in Table 2.3, entry 1). **Bottom.** Representative HPLC traces of the NCL of peptides **19** and **25** for reactions that proceeded for 0, 6 and 48 h. 220 nm channel (Blue), 270 nm channel (Red). The appearance of a signal displaying the same t_R as electrophile **25** after 48 h of reaction is marked as †. The oxidation of nucleophilic fragment **19** is marked as *.

Table 2.3: NCL of fragments **19** and **25**, and conditions screened for this coupling.



Entry	Temperature	Reductive Additive	Peptide conc.	pH	Detection of 18[a]	
					MALDI-TOF	HPLC
1	rt	10 eq TCEP	0.5 mM	7.1	Low intensity signal	traces
2[b]	rt	5 eq TCEP	1.0 mM	7.1	Not observed	Not observed
3	rt	5 eq TCEP	1.0 mM	6.8	Low intensity signal	traces
4[b]	rt	5 eq TCEP	1.0 mM	7.5	Low intensity signal	Not observed
5	40 °C	5 eq TCEP	0.5 mM	6.8	Low intensity signal	traces
6[b]	40 °C	5 eq TCEP	0.5 mM	7.1	Low intensity signal	traces
7[b]	40 °C	5 eq TCEP	0.5 mM	7.5	Low intensity signal	traces
8[c]	rt	-	0.5 mM	6.8	Not observed	Not observed
9[c]	rt	-	0.5 mM	7.1	Not observed	Not observed
10[c]	40 °C	-	0.5 mM	6.8	Not observed	Not observed
11[c]	40 °C	-	0.5 mM	7.1	Not observed	Not observed
12[c, d]	rt	-	0.5 mM	7.1	Not observed	Not observed
13[c]	rt	10 eq MPAA	0.5 mM	7.1	Not observed	Not observed
14	rt	40 eq MPAA	0.5 mM	7.1	Low intensity signal	traces

[a] The reaction mixture was monitored at 0, 6, 24 and 48 h. **18** detected as $[M+Na]^+$ = 1752.2 *m/z*.
[b] An intense signal corresponding to the oxidation of **19** was observed after 48 h.
[c] The oxidation of **19** was detected from 0 h.
[d] The buffer solution was de-gassed by freeze-pump thaw cycles.

2.3.5. Native chemical ligation using electrophiles 22 and 23

After these unsuccessful attempts at using thioester **25** as substrate for NCL, we set out to try peptide thioesters **22** and **23** which although reported to react relatively slower might not be affected by TCEP·HCl. Several couplings were carried using benzyl thioester **22**, but the peptide was found to have very limited solubility under the NCL conditions (the peptide was visibly suspended in the buffer even when ACN was used as co-solvent) and the ligation product was again detected only in trace amounts. Therefore, we focused on testing the MESNA derivative **23** instead.

However, upon scaling up the synthesis of electrophile **23** (Table 2.1, entry 2) for its study under NCL, two compounds sharing the mass of **23** ($[M-H]^-$ = 978.9 *m/z*) were detected in the reaction mixture by analytical HPLC (60:40 ratio; Figure 2.7). Tandem mass spectroscopy (MS/MS) showed that both compounds presented the same pattern of fragmentation masses implying they had the same sequence of amino acids (Figure 2.8). ^{1}H NMR spectroscopy analysis showed a great similarity between the two compounds. The most noticeable differences displayed by the signals corresponding to the C-terminal Ala residue (highlighted in Figure 2.9). These observations suggested that epimerization at the C-terminal amino acid could be occurring under the thioesterification conditions.

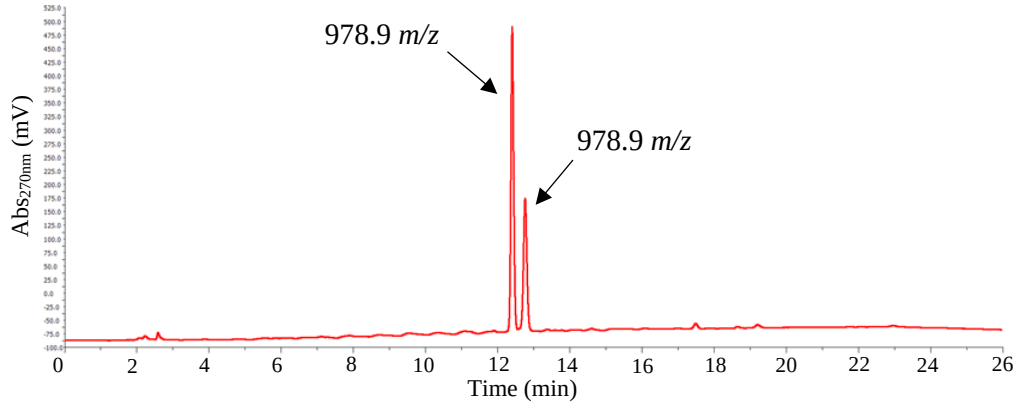


Figure 2.7: HPLC traces of the reaction mixture from the synthesis of thioester derivative **23** (reaction conditions in Table 2.1, entry 2).

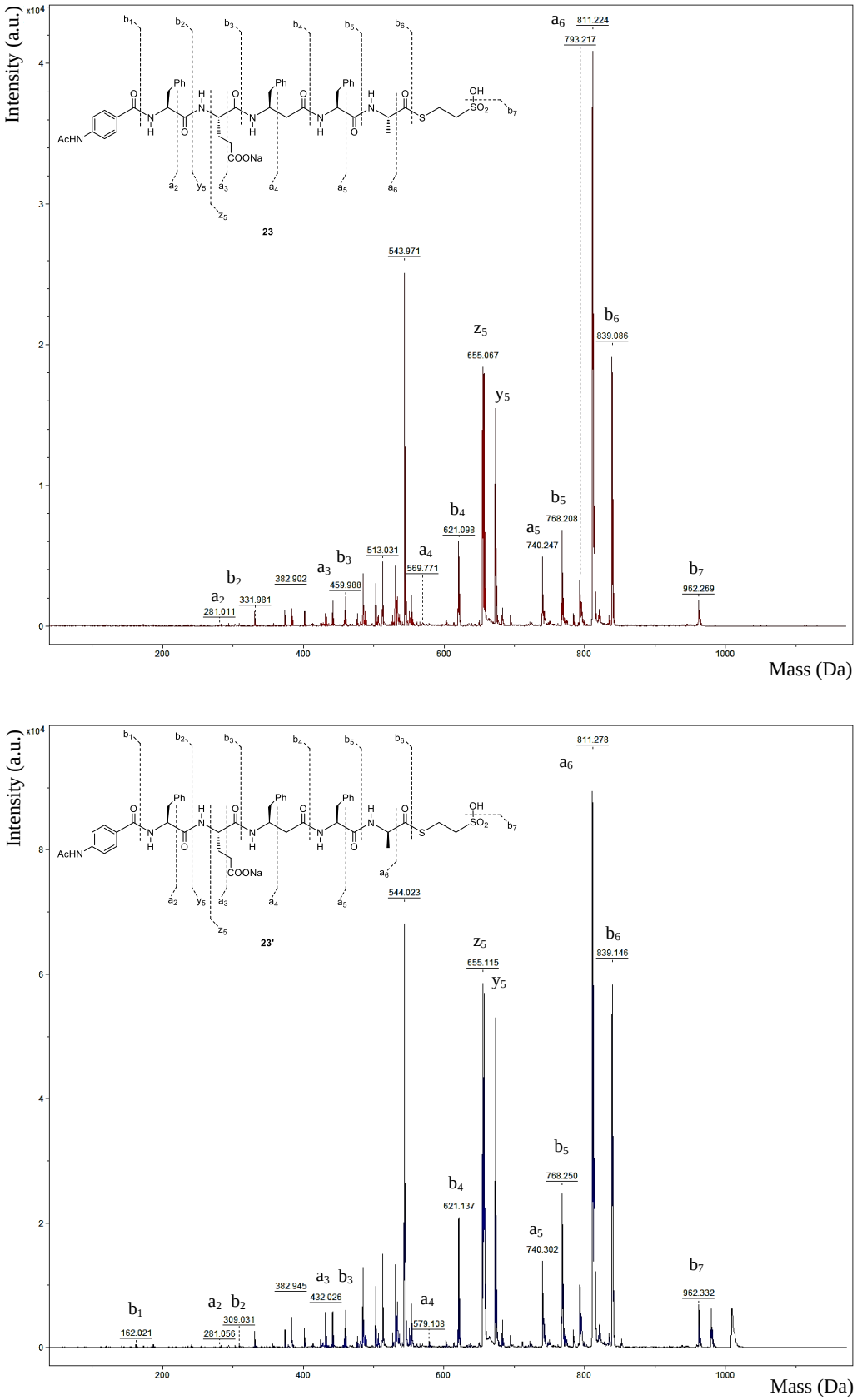


Figure 2.8: **Top.** Tandem mass spectra of thioester **23** ($[M+H]^+$ ion isotope-selected *m/z* = 980.0). **Bottom.** Tandem mass spectra of thioester **23'** ($[M+H]^+$ ion isotope-selected *m/z* = 980.0).

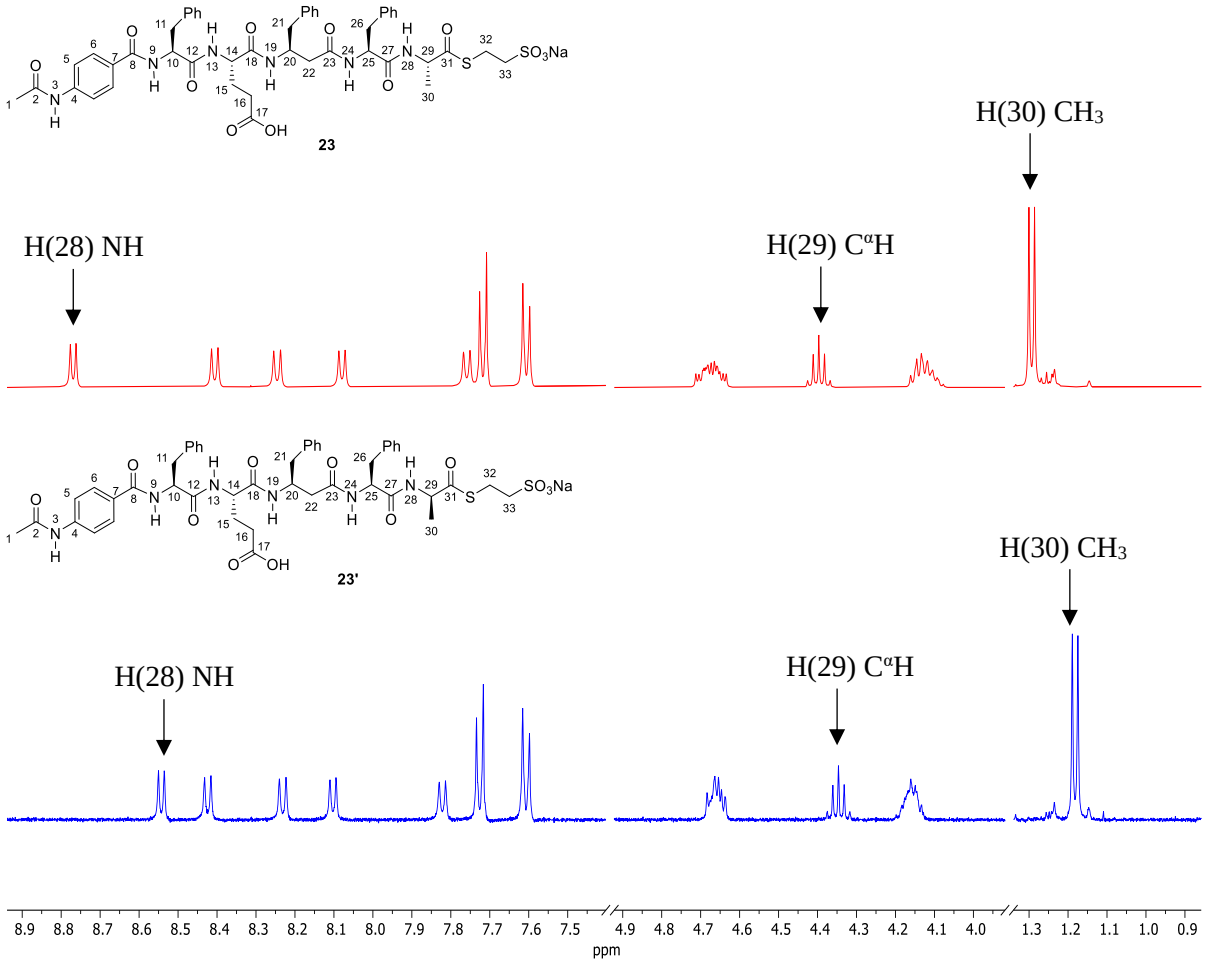
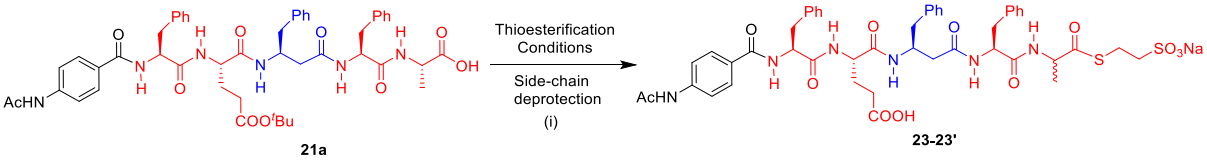


Figure 2.9: ^{1}H NMR spectra of the pair of peptide thioesters **23** and **23'**, the signals corresponding to the terminal Ala residues are shown by arrows.

To overcome this problem, reaction parameters including thioester equivalents used, reaction temperature and reaction sequence were modified (Table 2.4). It was found that when the reaction was carried out using a large excess of 2-mercaptoethanesulfonate (MESNA) at low temperature, thioester **23** was favoured (Table 2.4, entries 1 and 2). Conversely, the ratio of D-Ala epimer (**23'**) increased when peptide **21a** was treated with PyBop/DIPEA for longer reaction times before the addition of MESNA (Table 2.4, entry 3).

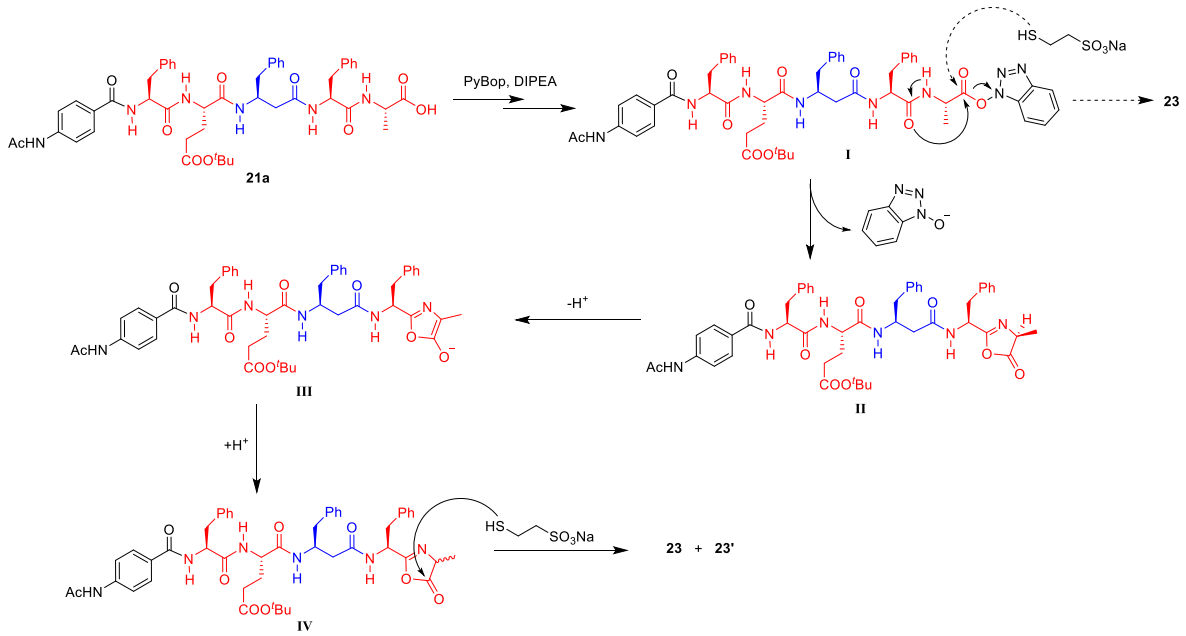
Table 2.4: Conditions screened for the synthesis of thioester **23**.



Entry	MESNA	PyBop/DIPEA	Rctn. time	Temperature	23 to 23' ratio[a]
1	20 eq	3 eq	45 min	0 °C	90:10
2	20 eq	3 eq	30 min	-20 °C	95:5
3[b]	5 eq	5 eq	2.5 h	rt	43:57

(i) TIPS/H_2O/TFA (2.5:2.5:95), 2 h.
[a] Calculated from analytical HPLC.
[b] Pre-treated with PyBop/DIPEA for 1.5 h.

It was rationalised that epimerization at the Ala residue could be happening *via* the mechanism proposed in Scheme 2.5. Upon generating the activated ester species **I**, intramolecular cyclization can occur to generate oxazolone intermediate **II**, which then can be deprotonated, epimerizing the $C^{\alpha}H$ centre (intermediate **III**).[87] Thus, immediate coupling with excess MESNA is essential for decreasing C-terminal racemization, whereas a prolonged activation with PyBop/DIPEA results in an increased epimerization.

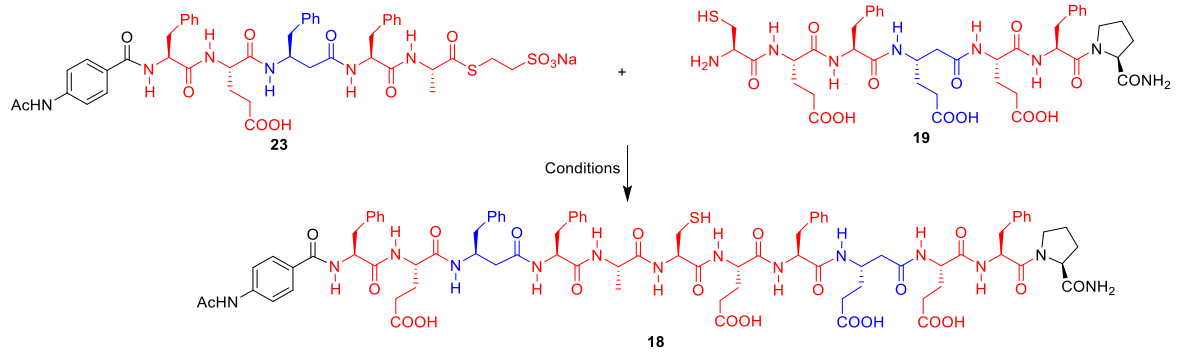


Scheme 2.5: Proposed mechanism for the generation of the mixture of peptide thioester **23** and **23'** epimers.

Once the synthesis of peptide thioester **23** had been optimised (Table 2.4, entry 2), we proceeded to try NCL with it (Table 2.5). This compound was shown to be more active than **22** and **25** counterparts; two

new peaks were observed in the 270 nm channel of the HPLC traces of the ligation reaction (Figure 2.10). Moreover, Analysis by MALDI-TOF MS revealed intense signals corresponding to the ligation product **18** and to the hydrolysis of the thioester **23**.[v] Addition of up to 12% ACN as co-solvent was found to improve the intensity of these signals (Table 2.5, entries 2 and 3).

Table 2.5: Conditions screened for the native chemical ligation using thioester **23**.



Entry	Temperature	ACN%	pH	Detection of 18[a]	
				MALDI-TOF	HPLC
1	rt	-	7.1	Strong signal	Low intensity peaks ~12.1 min
2	rt	12%	7.1	Strong signal	2 peaks ~12.1 min
3	40 °C	12%	7.1	Strong signal	2 peaks ~12.1 min

(i) 20 eq TCEP·HCl, peptides used at 0.5 mM.
[a] Reaction mixture was monitored at 0, 6, 24 and 48 h. **18** detected at $(M+Na)^+$ = 1752.2 *m/z*.

[v] This hydrolysis by-product of **23** (**26**, see Section 8.1.1) was generated separately by treating **21a** with the deprotecting mixture TFA/H_2O/TIPS 9.5:2.5:2.5 (v/v/v) for 2 hours. The free-carboxylic acid was then analysed by HPLC (t_R = 11.9 min) and MALDI-TOF (observed $[M-H]^-$ = 832.8 Da).

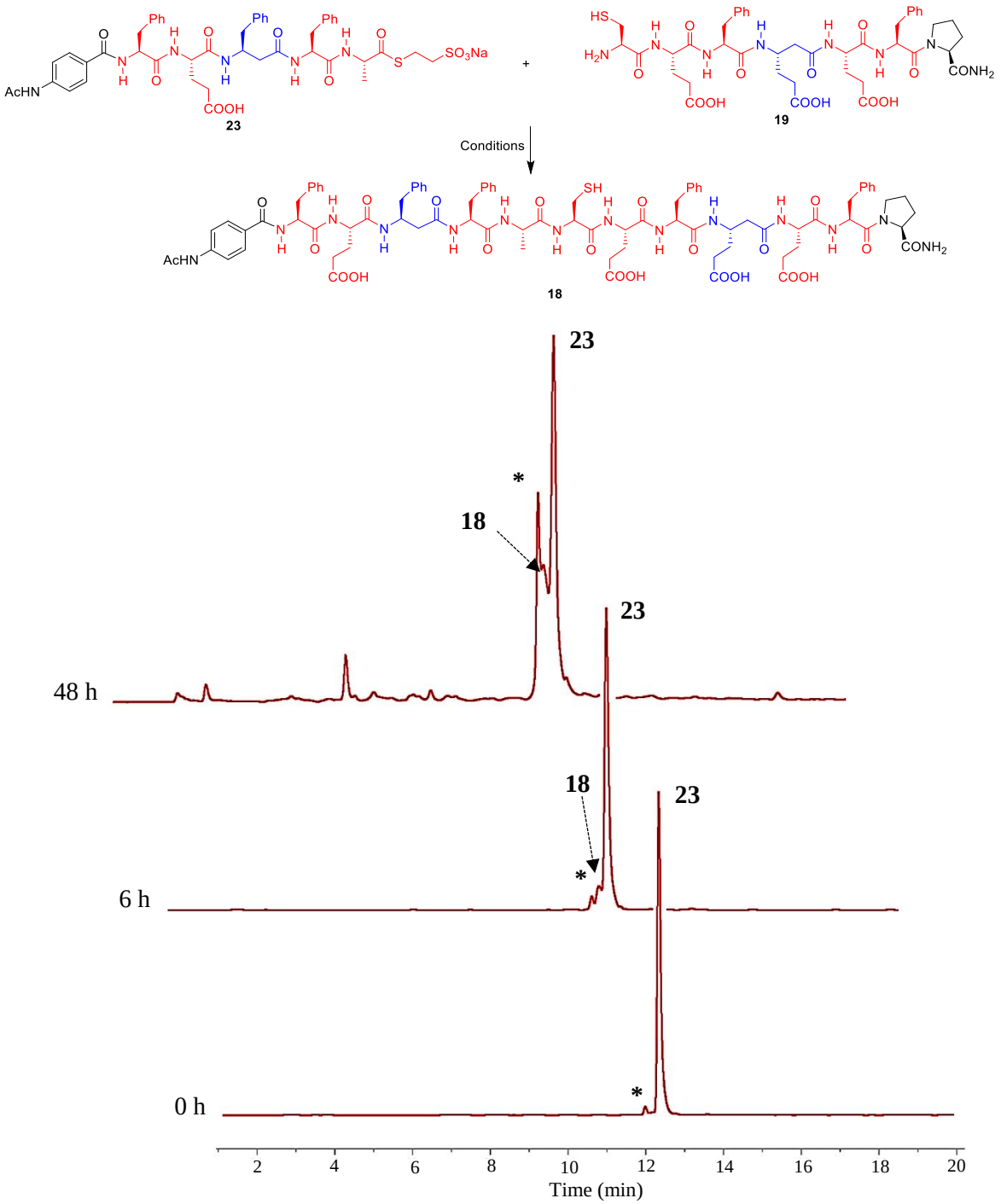


Figure 2.10: **Top**. Scheme of the NCL of peptides **19** and **23** (reaction conditions in table 4.5, entry 3). **Bottom.** Representative HPLC traces (270 nm channel) of the native chemical ligation of peptides **19** and **23** for reactions that proceeded for 0, 6 and 48 h. * denotes hydrolysis by-product **26** (see Section 8.1.1).

However, the identity of template **18** could not be fully confirmed by the analytical HPLC traces. Upon examination of the chromatogram it became clear that the method used so far was not suitable to resolve the mixture of **26**, **23**, and **18**, since their t_R were too similar, (Figure 2.10). Improving the chromatographic method to resolve the peptide mixture proved to be difficult. Longer runs combining isocratic and step gradients were attempted but were unsuccessful, generating complex chromatograms,

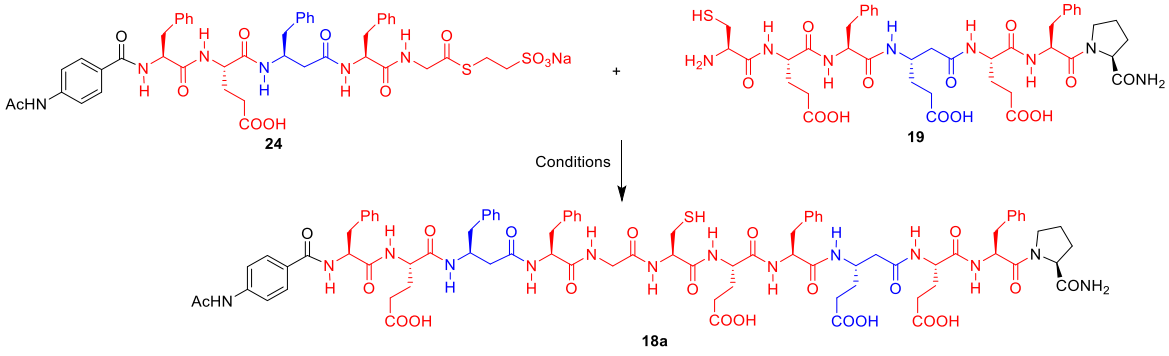
distorting peak shape or failing to completely resolve the mixture of compounds. This prevented quantification of the NCL product.

2.3.6. Native chemical ligation using Gly-derivative 24

So far, we considered that the efficiency of the electrophilic fragment was not adequate and needed to be improved. Modifying our original design by replacing the C-terminal Ala residue with a Gly residue in the electrophilic fragment (**21b**, Figure 2.5 and Scheme 2.4) could improve the peptide's reactivity.[88] This modification would also prevent racemization during the thioester synthesis. Thus, the MESNA derivative **24** was synthesised, and coupled to nucleophilic fragment **19** to generate **18a** (Table 2.6).

The electrophile was found to be considerably reactive. A larger percentage of ACN was needed as co-solvent in relation to the reactions involving Ala counterpart **23** (Table 2.6, entry 2). MALDI-TOF MS analysis showed **18a** and **27**[vi] (Gly analogues of **18** and **26**) were present in the reaction mixture (detected at $(M+Na)^+$ = 1738.7 Da and $(M-H)^-$ = 819.8 Da respectively). Unfortunately, this variant also led to an unresolved mixture of peptides in the HPLC traces. Longer chromatographic runs that incorporated isocratic elution and step gradients were attempted, but only partial peak-base separation was achieved (Figure 2.11). This prevented (again) accurate quantification of the NCL product.

Table 2.6: Conditions for the NCL of fragments **19** and **24** to generate Gly- template **18a**.



Entry	ACN%	pH	Detection of 18a[a]	
			MALDI-TOF	HPLC
1	12%	7.1	Strong signal	Small peak ~12 min
2	38%	7.1	Strong signal	Broad peak ~12 min

(i) 20 eq TCEP·HCl, peptides at 0.5 mM, rt [a] Reaction was monitored at 0, 6, 24 and 48 h.

[vi] Hydrolysis of thioester **24** was carried out in a similar fashion to that of **23**, generating free-carboxylic acid **27** (see Section 8.1.1) t_R = 11.9 min, $[M-H]^-$ = 819.8 Da (observed)

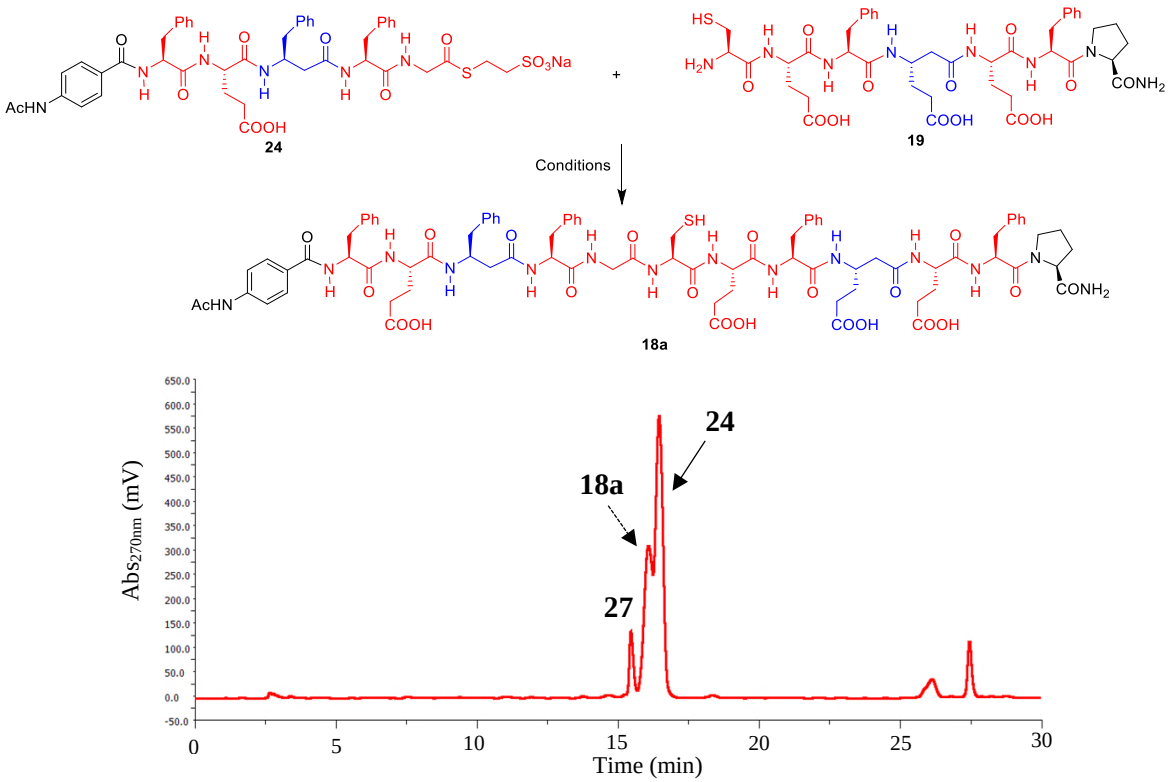


Figure 2.11: **Top**. Scheme of the NCL of peptides **19** and **24** (reaction conditions in table 4.6, entry 2). **Bottom.** HPLC traces of the NCL of peptides **24** and **19** after 48 h of reaction time. Chromatographic gradient:10 to 28% ACN in 0.1 M NH_4HCO_3 (aq) over 10 minutes; 28% ACN in 0.1 M NH_4HCO_3 (aq) for 10 minutes, 28% to 60 % ACN in 0.1 M NH_4HCO_3 (aq) over 20 minutes.

2.4. Reproducing the Ashkenasy system

2.4.1. Synthesis of the Ashkenasy replicator

During our experiments with the α/β-peptide variants of the Ashkenasy replicator, it was decided that the prototype system should be reproduced. We rationalised that by studying the model peptides we would have a better understanding of the conditions needed to observe the reported template-assisted autocatalysis. Moreover, these molecules would also be used for the control cross-templating reactions, where templates and fragments with mismatching information (α/β-residue sequence) would be mixed, expecting NCL to occur in a non-catalytic fashion (Figure 2.12).

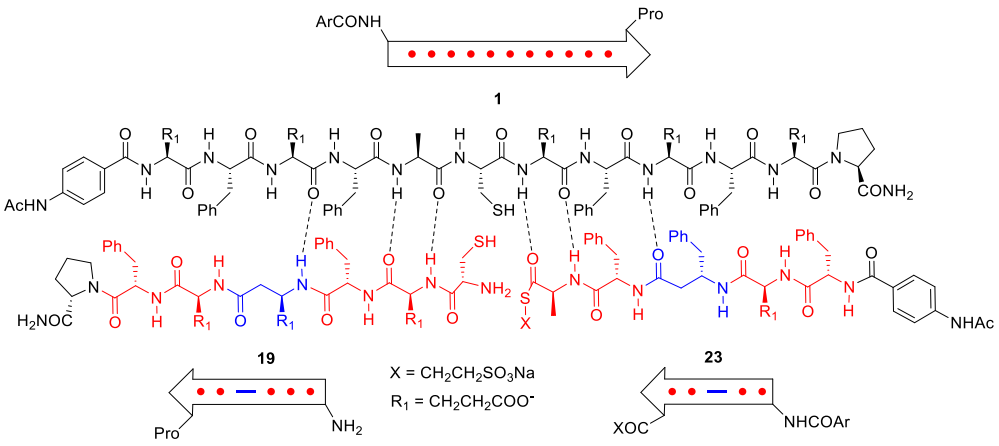
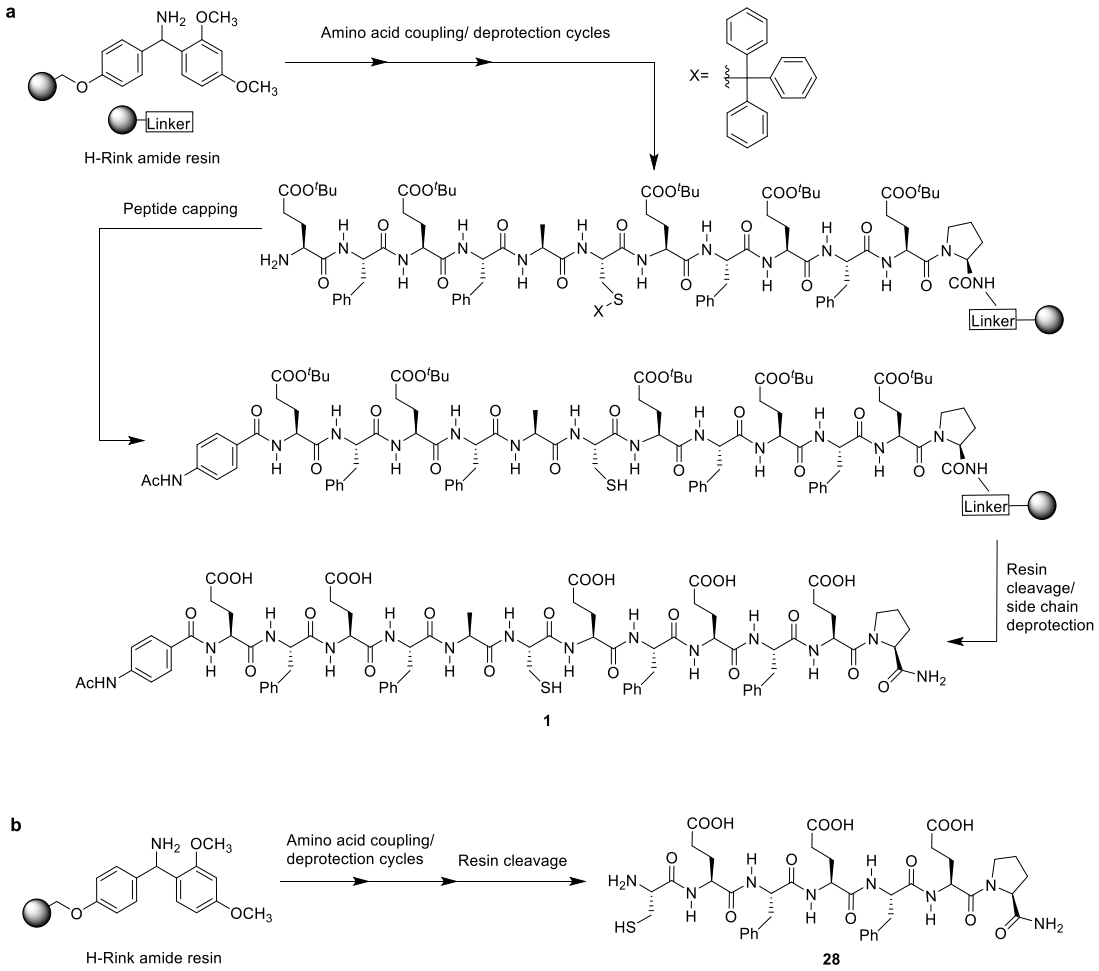
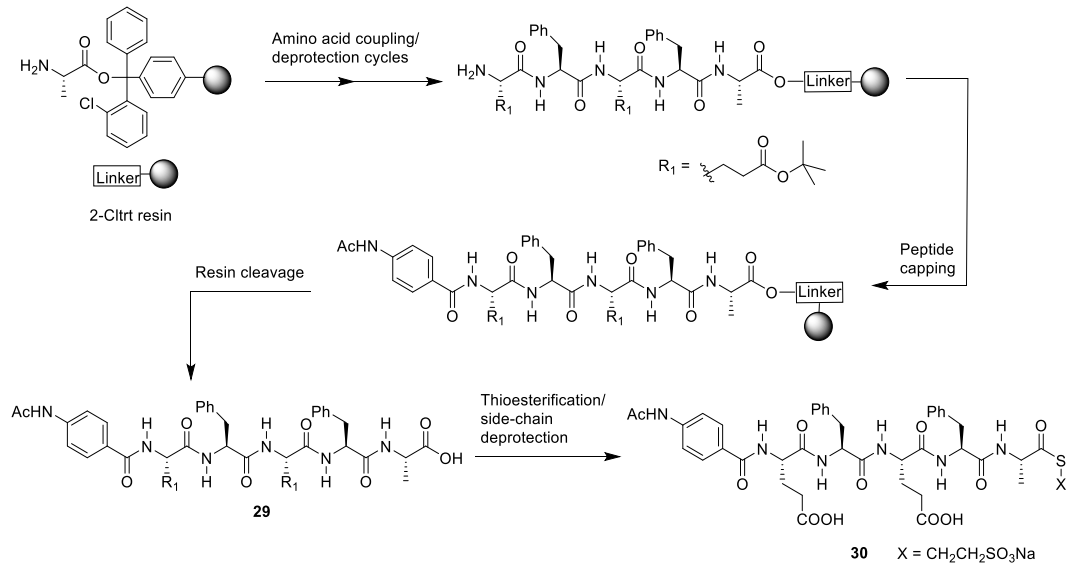


Figure 2.12: A control experiment consisting of the seeding of Ashkenasy's template **1** to the coupling of PMC-fragments **19** and **23**. The mismatching arrangement of residues and H-bonding is shown.

Thus, peptides **1**, **28** (labelled as **N** in section 1.4.2) and free-COOH peptide **29** were synthesised using Fmoc-based manual SPPS (Schemes 2.6 and 2.7). Once fragment **29** was obtained, the carboxylic acid at the C-terminus was transformed into a thioester (**30**) following the same methodology used in the synthesis of peptide **23**. The MESNA derivative was chosen given the reactivity it had displayed so far (*vide supra*) and its close similarity to the 3-mercaptopropionamide thioester used in the model peptide. The peptides were purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)). The fragments and template were characterised by analytical HPLC and MALDI-TOF MS.



Scheme 2.6: Overview of the SPPS (see section 7.2.1) of peptides **1** (**a**) and **28** (**b**).

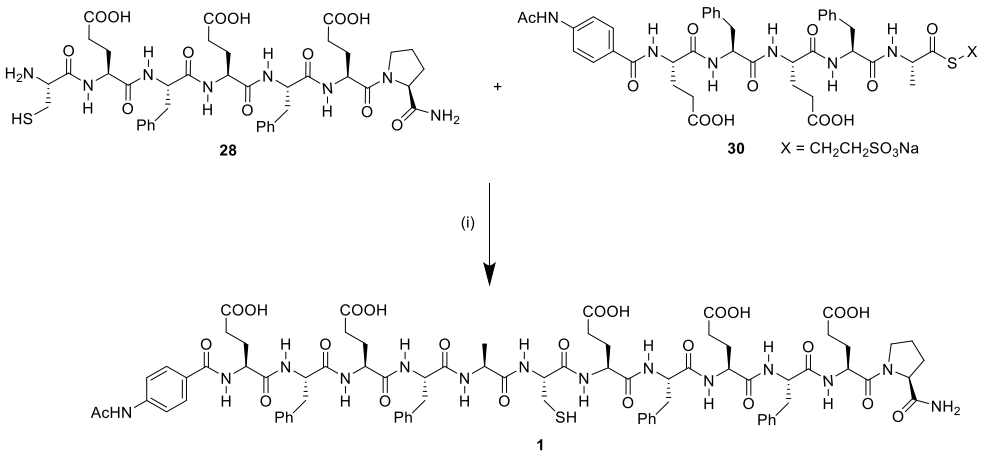


Scheme 2.7: Overview of the SPPS (see section 7.2.1) and thioesterification (see section 7.3.1) of peptide **29** to obtain **30**.

2.4.2. Evaluation of the templating effect of the Ashkenasy replicator

Relying on the MESNA-derived thioester **30** to be suitable for NCL, we set out to evaluate the self-replication of template **1**. Instead of reproducing the template-free couplings that had been attempted on our α/β-variant, we decided to focus on the template-seeded reactions (see section 1.4.2). We envisaged that since the seeded experiments exhibited the most pronounced accelerations of the coupling rates, reproducing these experiments presented the best opportunity to observe such behaviour.

Hence, the NCL of fragments **28** and **30** in the presence of template **1** was studied (Scheme 2.8). To do so, an analytical chromatographic method that achieved peak-base separation of the peptide signals was successfully developed (Figure 2.13). Moreover, emulating the methodology reported by Ashkenasy and co-workers,[48] the concentration of **1** in the reaction mixture was calculated using a foreign ABA-labelled peptide as standard (**Std**).



Scheme 2.8: NCL of peptide fragments **28** and **30** seeded with template **1**. (i) **1** (100 μM), TCEP·HCl (5mM) in MOPS buffer pH = 6.8-7.1, rt.

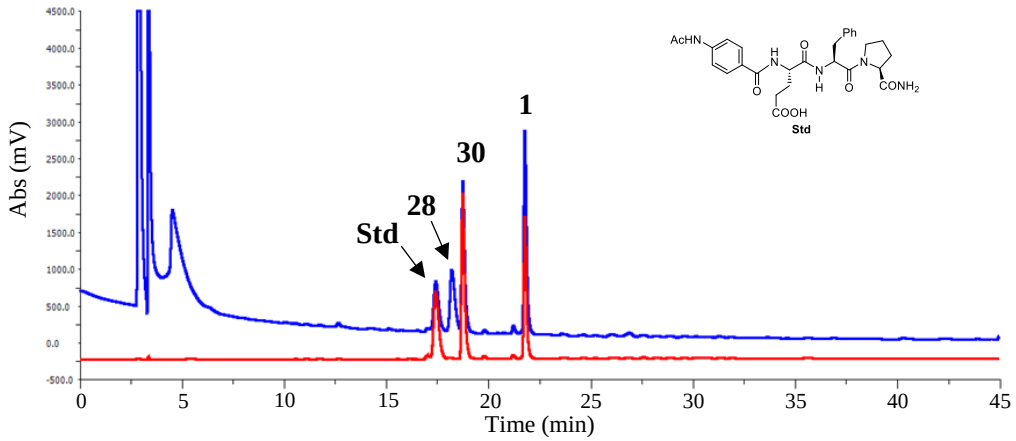


Figure 2.13: Representative HPLC traces of the NCL of peptide fragments **28** and **30** in the presence of template **1**. (reaction conditions Scheme 2.8). Gradient: 0 to 50% ACN in 0.1 M NH_4HCO_3 over 40 min. 220 nm channel (blue), 270 nm channel (red). The structure embedded in the traces corresponds to peptide standard **Std**.

Independent studies in the absence of fragments **28** and **30** were carried out to assess the accuracy of our analytical method to quantify **1** by comparison with the **Std** standard. It was found that the seeded amount of template (at ~100 µM concentrations) could be recovered with a standard deviation in the order of ±3% (for full details see section 7.4.3).

Ashkenasy and co-workers demonstrated the template-directed synthesis of peptide **1** by performing ligations seeded with 9 to 116 µM amounts of template. Since the highest initial concentrations of **1** generated the fastest ligations we set out to try NCL of peptides **28** and **30** in the presence of 100 µM of template.

Figure 2.14 shows representative chromatographic traces corresponding to samples taken at 0, 1 and 3 hours after the start of the reaction (a detailed description of the experimental procedure is outlined in section 7.4.2). An increase in the intensity of the peak corresponding to peptide **1** is noticeable as the reaction progresses, at the same time the signal corresponding to fragment **28** became less intense.

In order to follow the production of peptide **1** over time and evaluate the initial rates of ligation (first 180 minutes, see section 1.4.2), we registered the concentration of **1** in the reaction mixture at 8 different time points; most of them during the first hour at 0, 15, 30, 45 and 60 minutes after the ligation was initiated. To do this, some samples were analysed immediately, and others were stored frozen immediately after being taken to be analysed later.

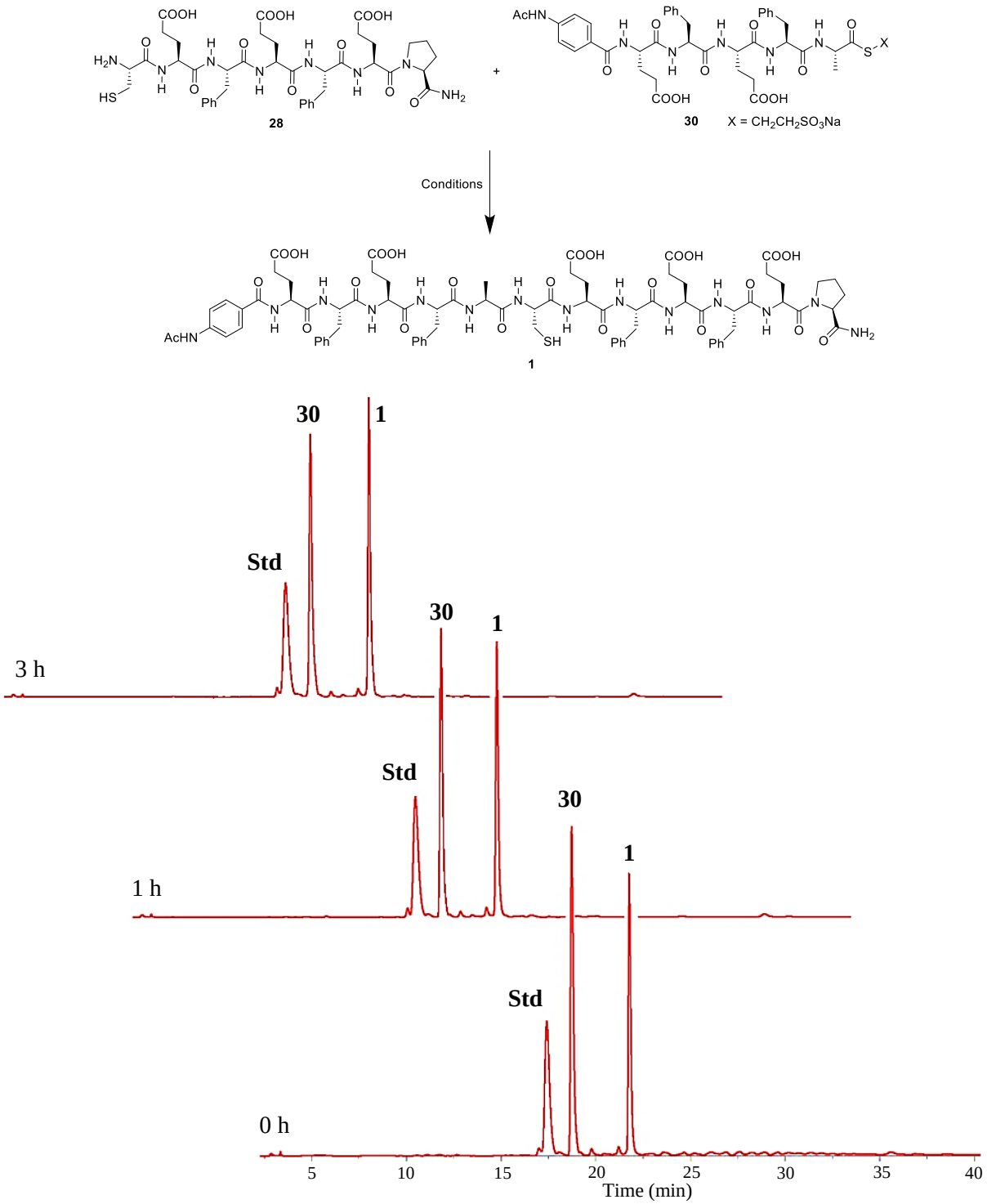


Figure 2.14: **Top**. Scheme of the NCL of peptides **28** and **30** (reaction conditions in Scheme 2.8, final pH =7.1). **Bottom.** Representative HPLC chromatograms (270 nm channel) of the NCL of peptides **28** and **30** in the presence of template **1** (100 µM) for reactions that proceeded for 0, 60 and 180 minutes.

Using this data, curves plotting the change in the concentration of template **1** throughout the reaction were built. Considering that the description given in the literature is ambiguous as to what the final pH of the ligations is, two sets of experiments were attempted, one with a final pH at 6.8 and a second with a final pH slightly above 7.0 (Figure 2.15). The graph in Figure 2.15a shows that no significant increase

in the amount of peptide **1** was produced during the first 180 minutes of the reactions with pH 6.8. But a trend pointing to a slow increase in the concentration of **1** (~20 µM) after 180 minutes of reaction was observed for the reaction with pH above 7.0. However, this apparent increase does not match the production of peptide **1** reported using a seed of ~100 µM of the template, and disappointingly the autocatalytic acceleration expected at this high concentration of seeding was not observed (see Figure 1.9 in section 1.4.2).

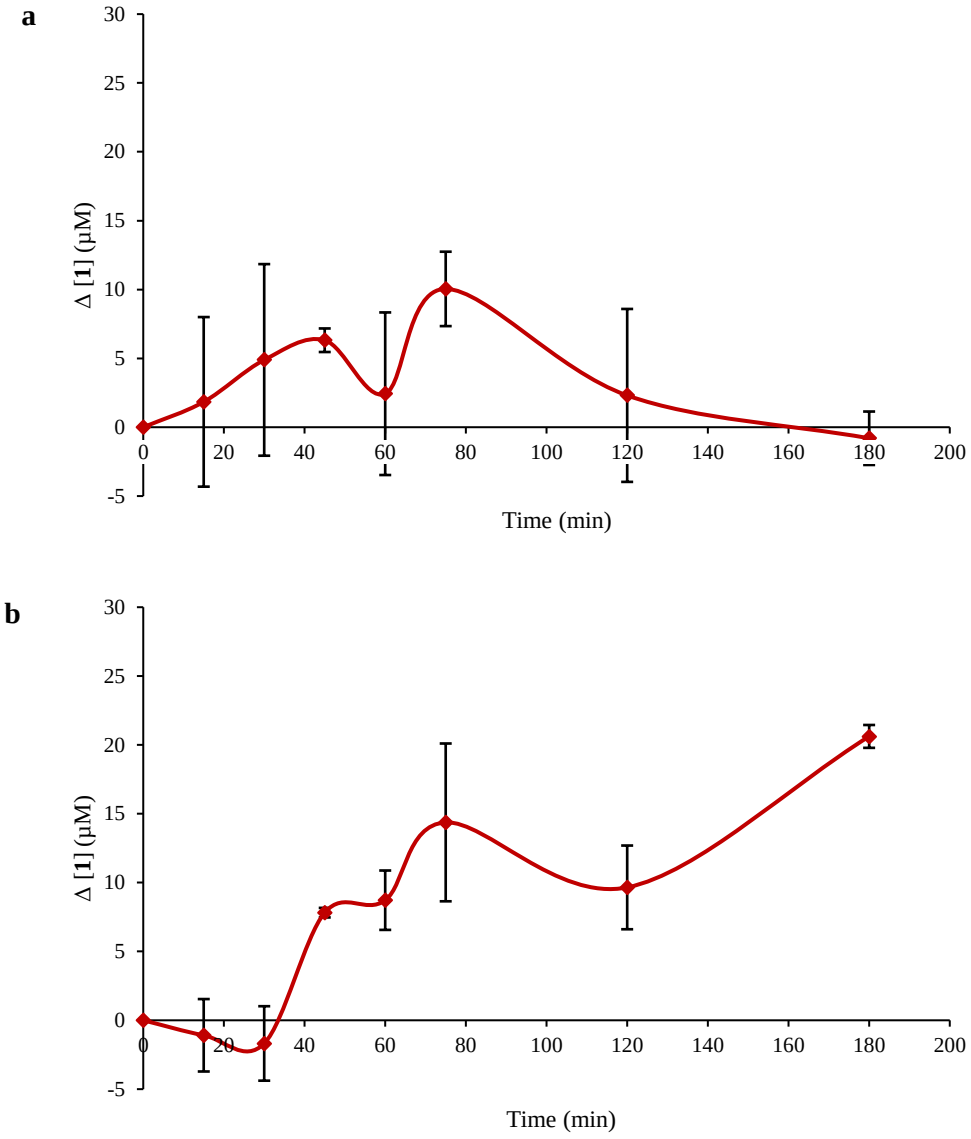


Figure 2.15: Production of peptide **1** over time in the NCL of fragments **28** and **30** seeded with 100 µM **1**. **a** MOPS buffer final pH = 6.8. **b** MOPS buffer final pH = 7.1.

At this point, we were worried about our sample treatment interfering with the quantification of **1**, so a series of ligations were monitored avoiding filtration of the samples prior to injection into the HPLC. Instead, centrifugation of the sample was used.[49] These reactions generated the curve shown in Figure 2.16. This graph shows that an increase of ~25 µM in the concentration of peptide **1** was produced after

120 minutes of reaction. Although this represented a slight improvement, the conversion after 3 hours of ligation was closer to that reported originally by Ashkenasy when a 41 μM seed of **1** was used instead of what was reported for the 116 μM seed (see Figure 1.9 in section 1.4.2). More importantly, the initial enhancement of the coupling rate (exponential trend, first 60 minutes) was still absent. Instead, the ligation displayed a behaviour similar to a template-free reaction (parabolic trend), where the rate of ligation increases only after a lag phase during which template **1** is produced slowly (see Figure 1.9 in section 1.4.2). Given that the lag phase in the template-free reactions corresponds to the slow build-up of a critical concentration of template which triggers the autocatalytic rate acceleration, this behaviour would not be expected for reactions with high concentrations of template already present in the mixture.

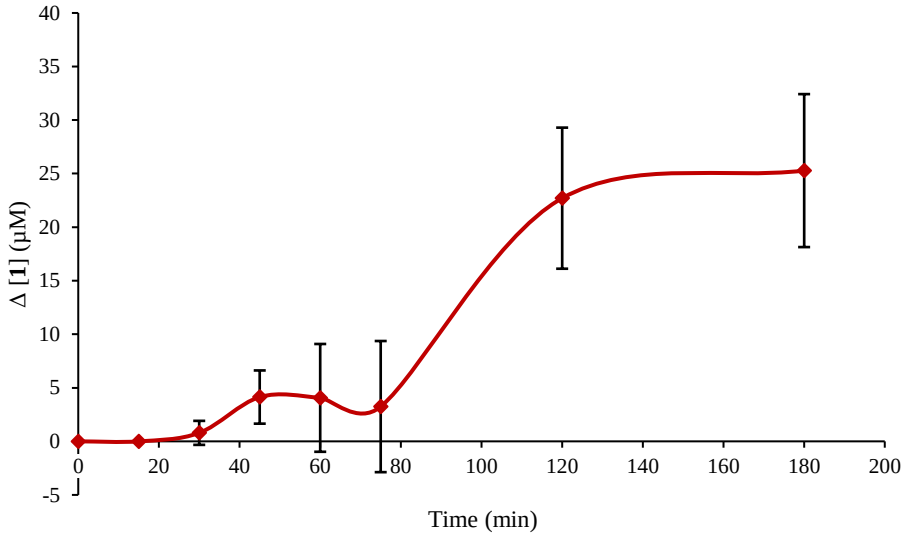


Figure 2.16: Production of peptide **1** over time in reactions between fragments **28** and **30** that were initially seeded with 100 μM of **1** as a template. MOPS buffer final pH = 7.1.

2.4.3. Brief structural study of peptide 1

These negative results prompted us to study the structural behaviour of peptide **1**. As it has been described in section 2.1, Ashkenasy and co-workers characterised the assembly of **1** into a variety of supramolecular structures and matched this process to the course of the replication experiments. This allowed them to determine the nature of the species responsible of the catalytic effect. Amongst the analytical techniques used to characterise such aggregates was dynamic light scattering (DLS, Figure 2.17, top). Measurements of diluted solutions of peptide **1** revealed the formation of mainly large aggregates, while measurements of more concentrated solutions revealed the formation of increasing amounts of smaller aggregates, the latter are hypothesised to be the species acting as active catalysts.

Carrying out an analogous experiment (Figure 2.17 bottom) it was found that at 100 μM only the larger non-catalytically active aggregates of **1** were present. This observation was correlated to the lack of self-replicating activity and pointed to a flaw in the experimental conditions, which was tampering with the assembly of the β-sheet nanostructures responsible for such activity.

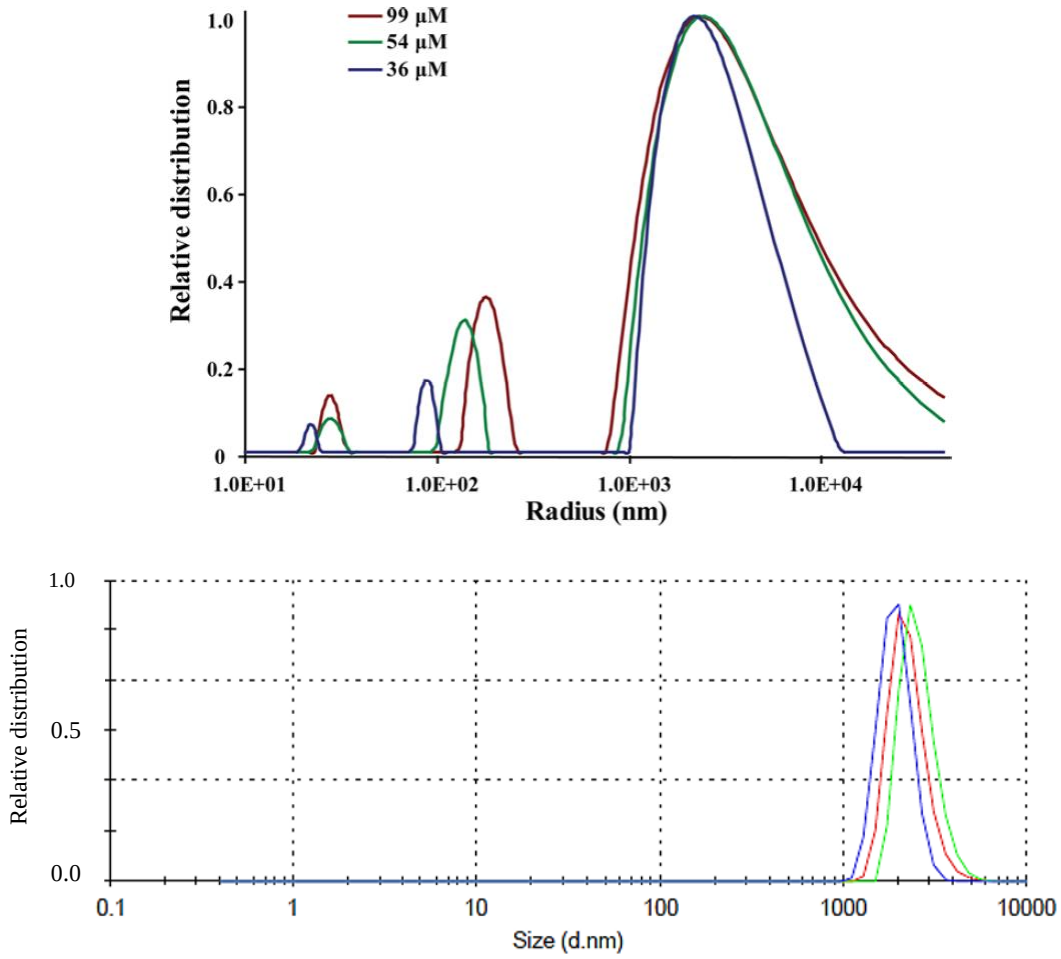


Figure 2.17: **Top**. Intensity weighted size distribution of peptide **1**, obtained from DLS experiments in buffered solutions (ammonium carbonate buffer pH 7.0) at various concentrations (taken from the supporting information provided for reference 48). **Bottom**. Intensity weighted size distribution of peptide **1** (100 μM) obtained from DLS experiments in buffered solutions (0.1 M ammonium carbonate buffer pH = 7.1) Z-Average size (d.nm): 3173, PDI: 0.353.

2.5. Conclusions and future work

2.5.1. Amphiphilic PMC system

Novel α/β-peptides **22-25** were designed and synthesised based on a replicator reported in the literature. Our aim was to gain evidence of the PMC concept by showing we could achieve template-directed catalysis in the NCL of α/β-peptides fragments with matching sequences of α/β-residues. To do this, we attempted to couple fragment **19** with a series of thioester fragments through NCL. These couplings were unsuccessful using thioesters **22** and **25**, due to solubility problems and incompatibility with TCEP additive, respectively. More promising results were obtained by using MESNA-derived thioester **23** and its Gly analogue **24**, where strong evidence of the generation of the ligation products was obtained, both from MALDI-TOF MS and HPLC. However, the resultant mixture could not be resolved by analytical chromatographic methods, and therefore full identification (and quantification) of the ligation products was impossible.

Considering that the most significant issues encountered while attempting to analyse NCL reactions have their origin in the close similarity of the t_R of peptide thioesters and ligation products, we envisioned that modifying the polarity of the electrophilic thioester peptide fragment could help to differentiate it from the ligation product in the HPLC traces. This could be achieved by inverting the sequence of alternating hydrophobic and hydrophilic amino acids. Thus, by changing the Phe residue immediately adjacent to the C-terminal Pro of product **18** for a Glu residue and then building on the alternating polarity sequence, alternative peptide **31** would result. The peptide preserves the facial amphiphilicity and self-complementarity displayed by peptide **18** (Figure 2.18). However, fragments **32** and **33** now have a better balance of hydrophobic Phe *vs* hydrophilic Glu residues, specially thioester fragment **33** (in comparison to thioester fragment **23**). We hypothesise that this modification should decrease the retention time of fragment **33** differentiating it from peptide **31**. If that is so we should be able to evaluate the rates of NCL and perhaps demonstrate the existence of an autocatalytic effect (however, see below).

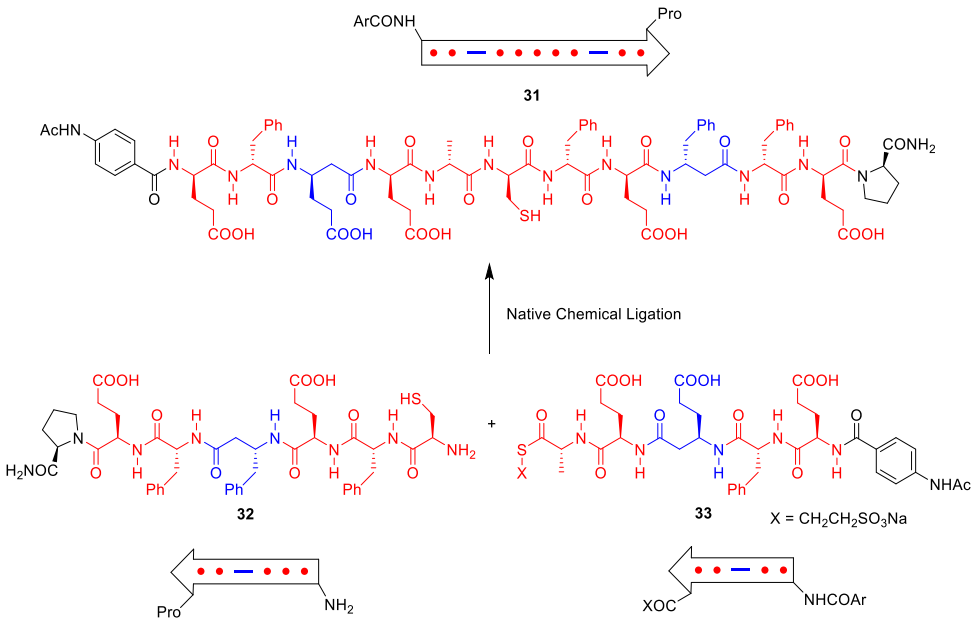


Figure 2.18: Alternative PMC system with an inverse sequence of amino acid polarity.

2.5.2. Ashkenasy's all-alpha system

The study of the system reported by Ashkenasy and co-workers revealed the many experimental variables that need to be optimised for the system to exhibit its auto-catalytic activity and for it to be detected analytically. So far, this has not yet been achieved. Seeding experiments have shown a lack of the initial ligation rate enhancement, and conversions of template **1** are below the original reports. On the other hand, key supramolecular assemblies have not been detected by DLS.

Although there is further work that could be done on this system, the difficulties experienced may suggest that Ashkenasy's approach is not, after all, ideal for demonstrating molecular Morse code. Firstly, the use of water as solvent constrains the structures and coupling methodology that can be employed. In particular, NCL is the only realistic coupling method, and this has disadvantages. The requirement for cysteine as the nucleophilic residue means that the system is vulnerable to oxidation, and the electrophilic components (thioesters) are not indefinitely stable to the aqueous solvent. Secondly kinetic studies on autocatalysis by molecules which self-assemble into ill-defined aggregates may always be difficult to reproduce. Catalytic efficiency will vary according to the aggregation state, and this will be influenced by many factors. Ashkenasy had no alternative to kinetic studies, but for molecular Morse code autocatalysis may in principle be detected through *selectivity*, and this could be more reliable and effective. For further development of this line of thought, see chapter 4 and 5.

Chapter 3. Influence of the α/β-residue sequence in the formation of β-hairpins

The work described in this chapter was designed by the author of this thesis (AAC) and Prof. Anthony Davis. Some of the work in this chapter was carried out in collaboration with Kelly Chu and has been reported in her thesis towards an MSci degree at the University of Bristol. Specific contributions are mentioned in the text and in sections 7 and 8. All other experiments and analyses were carried out by AAC.

3.1. Background

Balaram and co-workers have reported the design of short peptide β-hairpins (see section 1.6.1) with β-strands composed of α-,[89,90] β-[91] and a mixture of both α- and β-residues[75,92] (Figure 3.1). In all the examples, a β-turn-nucleating ^{D}Pro-Gly segment was used to promote the folding of the peptide strands. The conformation adopted by these molecules in solvents of medium polarity (methanol and chloroform) was assessed using CD and NMR, mainly through the observation of critical cross-strand NOEs. Figure 3.1 shows the conformation of hairpin **5** in CD_3OH. The schematic shows the ^{D}Pro-Gly segment, which adopts a type-II' β-turn conformation and five intramolecular cross-strand hydrogen bonds stabilizing the peptide fold. The β^3hPhe residues occupy facing positions on the hairpin, with the side chains projecting on opposite faces of the β-sheet. Side-chain groups were chosen to promote β-sheet formation and solubility in organic solvents (CH_2Ph, *i*Pr, *i*Bu).[92]

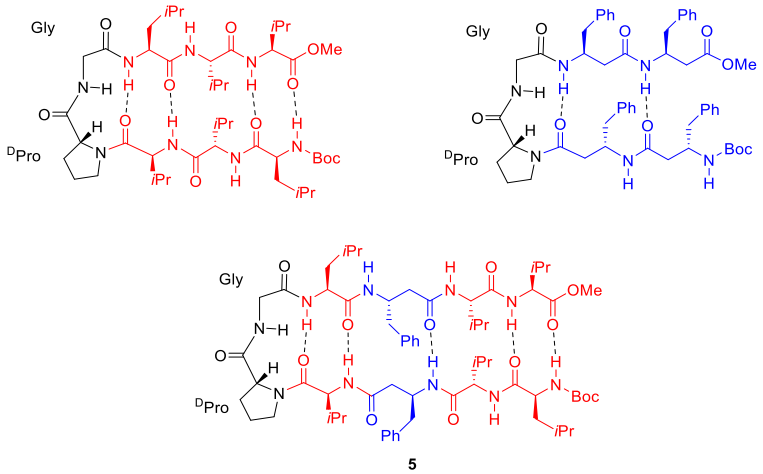


Figure 3.1: Short-chain peptides folded into the β-hairpin motif reported by Balaram and co-workers. **Top**. Examples containing all α (left) and all β (right) residues. **Bottom.** β-hairpin **5**, an example containing both, α- and β- residues.

Crystal structure analyses performed by Balaram and co-workers have confirmed that β-hairpin molecules constructed from α-, β- and α/β-amino acids have a strong propensity to assemble into regular, extended, infinite β-sheets by multiple and repeated hydrogen-bonding.

3.2. Project aims

The objective of this project was to provide a structural basis for peptide Morse code by investigating the propensity of α/β-peptides to adopt a β-sheet secondary structure depending on their matched *vs* mismatched sequences of α/β-residues. To do this, we used the β-hairpin motif, which promotes the formation of an intramolecular β-sheet array. We rationalised that this intramolecular folded structure would be simpler to study and control than it would be extended aggregates in water. Given that the PMC hypothesis centres in the transmission of information through H-bonding complementarity, we attempted to minimise other factors influencing molecular recognition by studying our molecules in organic media, where hydrophobic interactions would be weakened in favour of interactions based of H-bonding. To the same effect, we used amino acid residues bearing hydrophobic side chains only. Balaram's work provided a starting point, but only a few sequences had been explored by his group. Thus, we built on Balaram's literature work by examining the conformation in organic solvents adopted by peptides bearing a central Gly-^{D}Pro segment and strands of matching or mismatching α/β-sequences (Figure 3.2). We predicted that internally matched examples would be folded as stable hairpins whereas mismatched cases would adopt alternative conformations. This would demonstrate the sequence selective formation of intramolecular β-sheets and our ability to control peptide conformation through PMC.

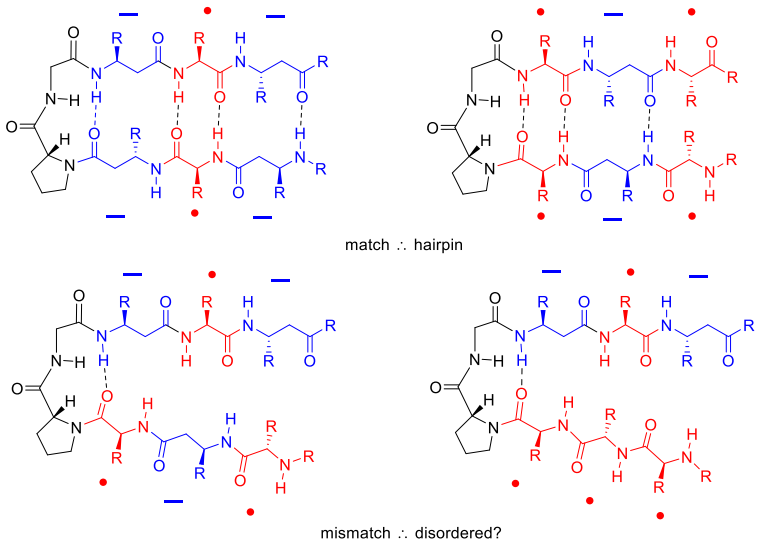


Figure 3.2: **Top**. Folded peptides containing a Gly-^{D}Pro nucleus with matching α/β-strands. **Bottom**. Folded peptides containing a Gly-^{D}Pro nucleus with α/β-mismatching strands.

3.3. A series of decamer α/β-peptide β-hairpins

3.3.1. Peptide design

To probe the assembly of the β-hairpin motif in peptides with matched *vs* mismatched α/β-residues, four decapeptides (Figure 3.3) based on Balaram's work were designed. Peptide **34** is based directly on Balaram's hairpin **5**, the only difference being the N-terminus protecting group, changed from Boc to Fmoc for synthetic reasons (we used Fmoc-based SPPS). The other matched sequence (**35**) is an isomer of **34** where the two facing β^3hPhe residues have been moved one position away from the β-turn. For the mismatches, we were interested in evaluating a variant where one of the β^3hPhe residues was substituted for a Phe residue, so that one of the strands contained all α-residues (**36**). On the other hand, in the more ambitious mismatched peptide **37** (an isomer of peptides **34** and **35)** the β-residues have been moved two residues away from each other in the strands. The balance between Leu and Val residues in the polypeptides intended to promote good signal dispersion in ^{1}H NMR.

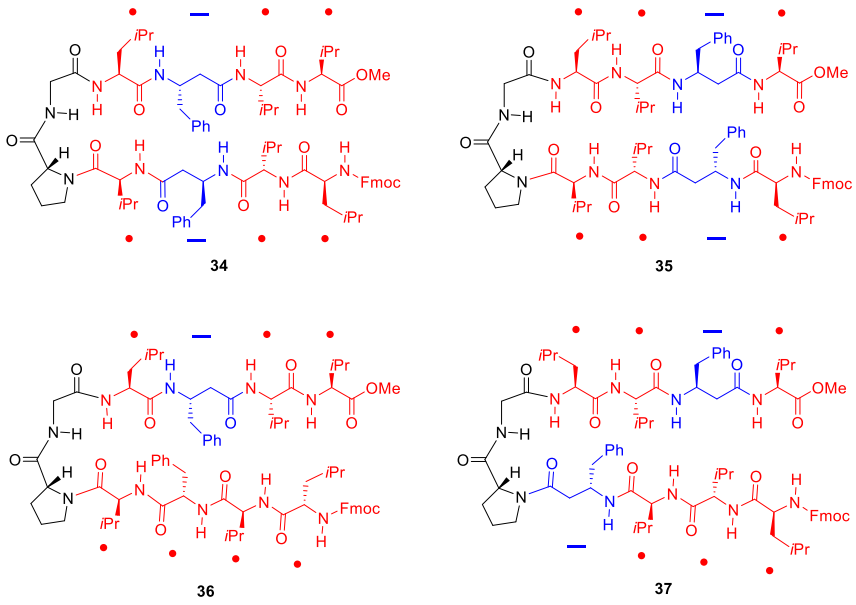
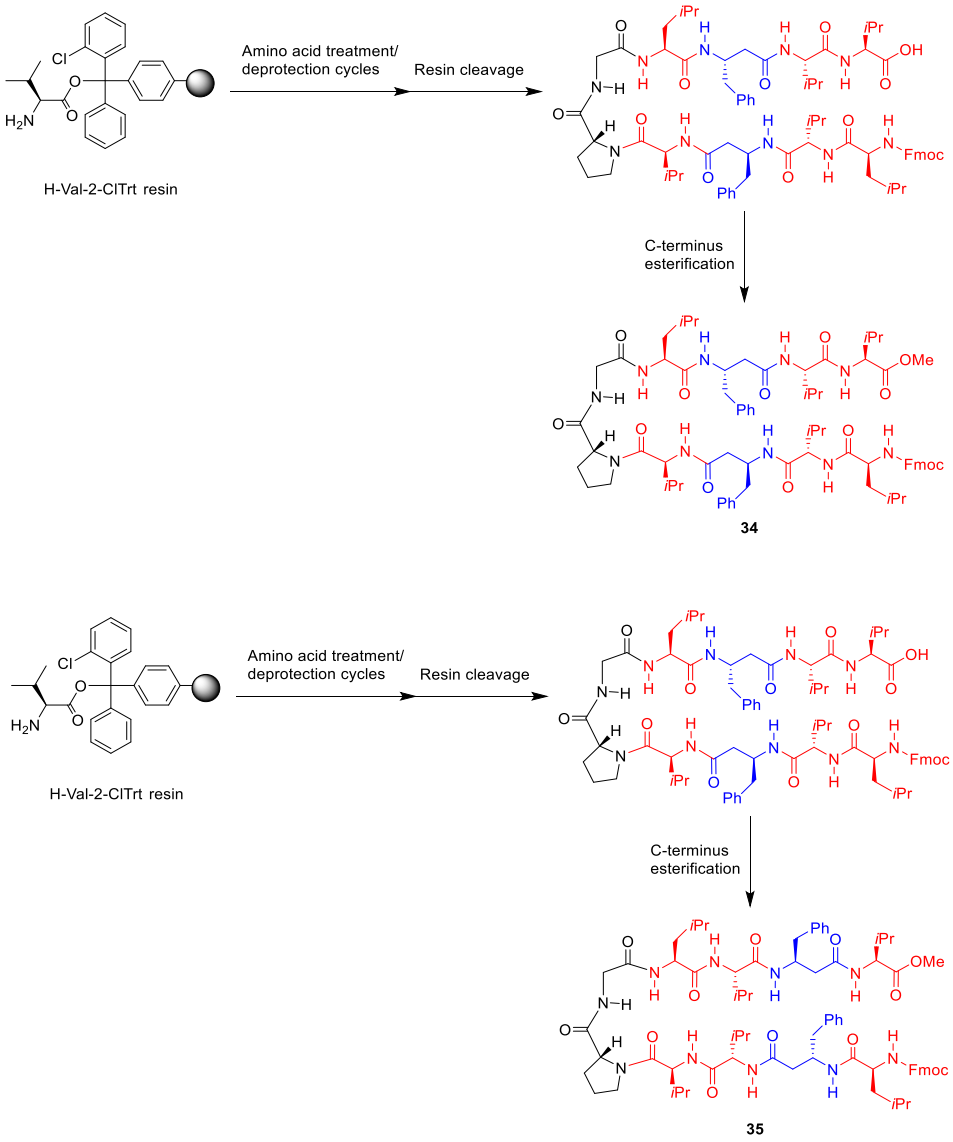


Figure 3.3: **Top**. Decapeptides bearing matching α/β-strands. **Bottom**. Decapeptides with mismatching strands.

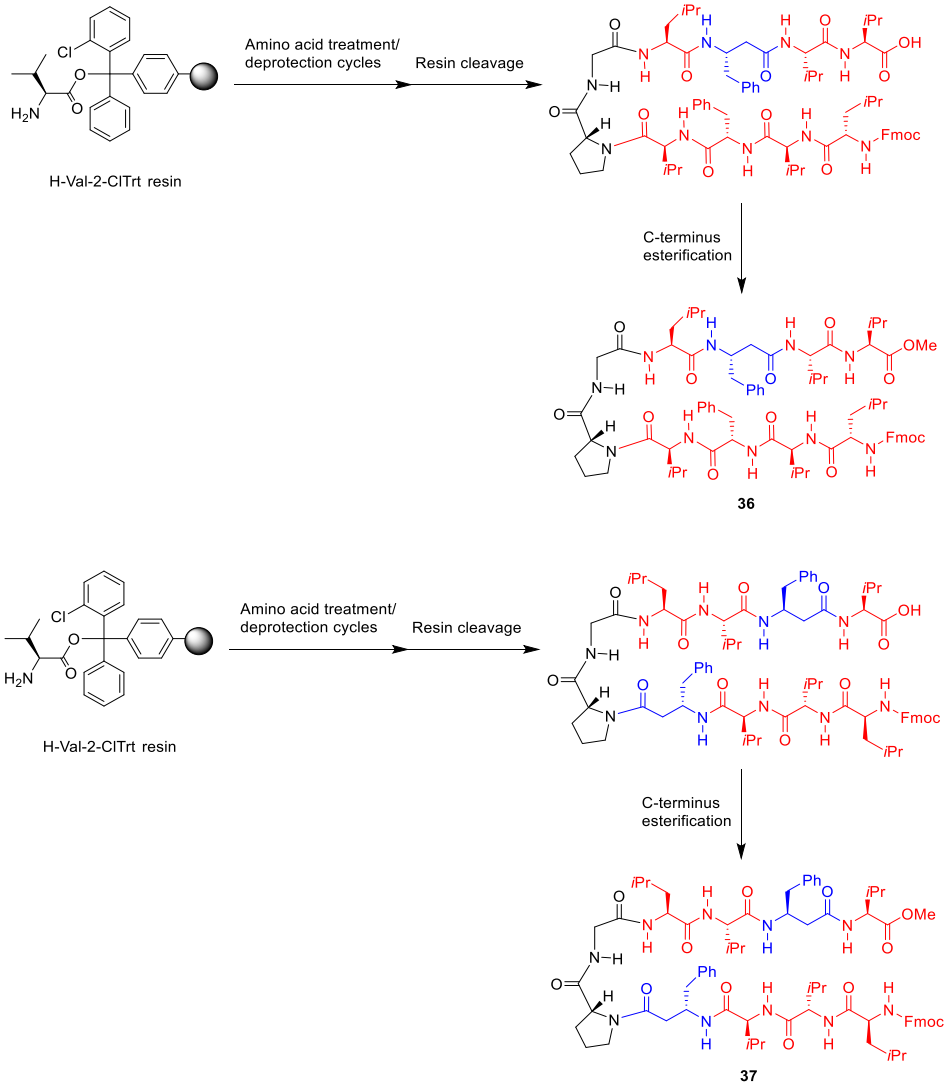
3.3.2. Peptide synthesis

The peptides were synthesised using a standard manual Fmoc-based SPPS methodology to yield the C-terminal free carboxylic acids. These were then converted into the appropriate methyl ester derivatives (Schemes 3.1 and 3.2). The Fmoc-deprotection step on the N-terminal residue was omitted. The peptides were purified through HPLC (methanol/water). The identity and purity of the peptides was

assessed by MS and analytical HPLC, and for a selection of peptides in this chapter ^{1}H and ^{13}C NMR characterization was also carried out (SPPS of peptides **34** and **36** was carried out by Kelly Chu).



Scheme 3.1: Overview of the synthesis of decamers **34** (**top**) and **35** (**bottom**; for full conditions see sections 7.2.1 and 7.3.2).



Scheme 3.2: Overview of the synthesis of decamers **36** (**top**) and **37** (**bottom**; for full conditions see sections 7.2.1 and 7.3.2).

3.3.3. Conformational studies of decamers 34 and 36 by NMR in CD_3OH

Given the close similarity of peptide **34** to literature compound **5**, we set out to study the conformation of the former using the analyses reported for **5** (see reference 77) under similar conditions. We expected our peptide to exhibit a very similar behaviour to that observed for Balaram's hairpin **5**. In the literature, **5** was studied in CD_3OH solutions at ~4 mM. However, peptide **34** could not be dissolved in CD_3OH to those concentrations, saturating beforehand. To overcome this, $(CD_3)_2SO$ was used as co-solvent. This difference in the peptide's solubility could only be attributed to the replacement of the N-terminus Boc protecting group in peptide **5** for an Fmoc protecting group in peptide **34**.

To study the conformation of the peptide in solution, a complete assignment of proton resonance signals was carried out. Residue-specific assignments were obtained from a combination of TOCSY, COSY, and HSQC NMR experiments, whereas ROESY and NOESY NMR spectroscopy permitted sequence-specific assignments.

A fully assigned ^{1}H NMR spectrum of peptide **34** in $CD_3OH/(CD_3)_2SO$ (3:1) is shown in Figure 3.4 (this assignment was carried out in collaboration with Kelly Chu). The spectrum shows a well dispersed group of NH resonances (ranging from δ = 7.3 to 8.6 ppm) with minimal overlap with the aromatic CH signals corresponding to the Fmoc protecting group. In the $C^{\alpha}H$ region, resonances are dispersed over the range of δ = 4.2 to 4.9 ppm, with some degree of overlap. The backbone NH groups with the exception of Leu1 (see Figure 3.4), which is at the N-terminus, all present $^3J_{NH\text{-}C\alpha H}$ values of $\delta \geq 8$ Hz, characteristic of extended β-sheet conformations.[93]

Once the ^{1}H NMR spectrum of **34** had been assigned, we attempted to evaluate the peptide's conformation in solution by looking for characteristic cross-strand contacts. Figure 3.5 shows a partial ROESY NMR spectrum of peptide **34**. A single cross-strand NOE was observed (β^3hPhe3 $C^{\alpha}H \leftrightarrow \beta^3$hPhe8 $C^{\beta}H$, anticipated in an intramolecular antiparallel sheet). Furthermore, sequential ^{D}Pro5 $C^{\delta}H \leftrightarrow$ Val4 $C^{\alpha}H$, and (relatively weaker) Gly6↔Leu7 NH↔NH (d_{NN}) NOE interactions are consistent with a type II' β-turn at the ^{D}Pro-Gly segment.[71] The Gly6↔Leu7 contact is somewhat ambiguous since the amide Gly6 and Val10 signals overlap in the ^{1}H NMR spectrum. However, a Val10↔Leu7 (d_{NN}, *i*/*i* + 3) contact would be unexpected for a common ordered secondary structure.

Unfortunately, we found no evidence for other expected cross-strand NOEs; Val4↔Leu7 (d_{NN}), Leu1↔Val10 (d_{NN}) and Val2 $C^{\alpha}H \leftrightarrow$ Val9 $C^{\alpha}H$ (all reported for peptide **5**).[92] In the case of Val2, the corresponding $C^{\alpha}H$ signal overlaps with the residual water peak and any NOE falls under the noise and cannot be observed (Figure 3.5, the $C^{\alpha}H$ signal was assigned using HSQC NMR).

The absence of these cross-strand NOEs suggested that $(CD_3)_2SO$ could be disrupting the assembly of peptide **34** into a β-hairpin. Considering DMSO is a strong H-bond acceptor, it is possible that it could

be solvating the peptide backbone NH groups and preventing intramolecular association from occurring.[94] Thus, both NOESY and ROESY NMR spectra were recorded in a 600 MHz spectrometer at 25 and 15 °C of **34** in CD_3OH (~1 mM). Cross-strand NOE (β^3hPhe3 $C^{\alpha}H \leftrightarrow \beta^3$hPhe8 $C^{\beta}H$ was observed again but disappointingly, the missing NH↔NH NOEs could not be observed, and in the case of Val2, the residual solvent signal continued to be an issue.

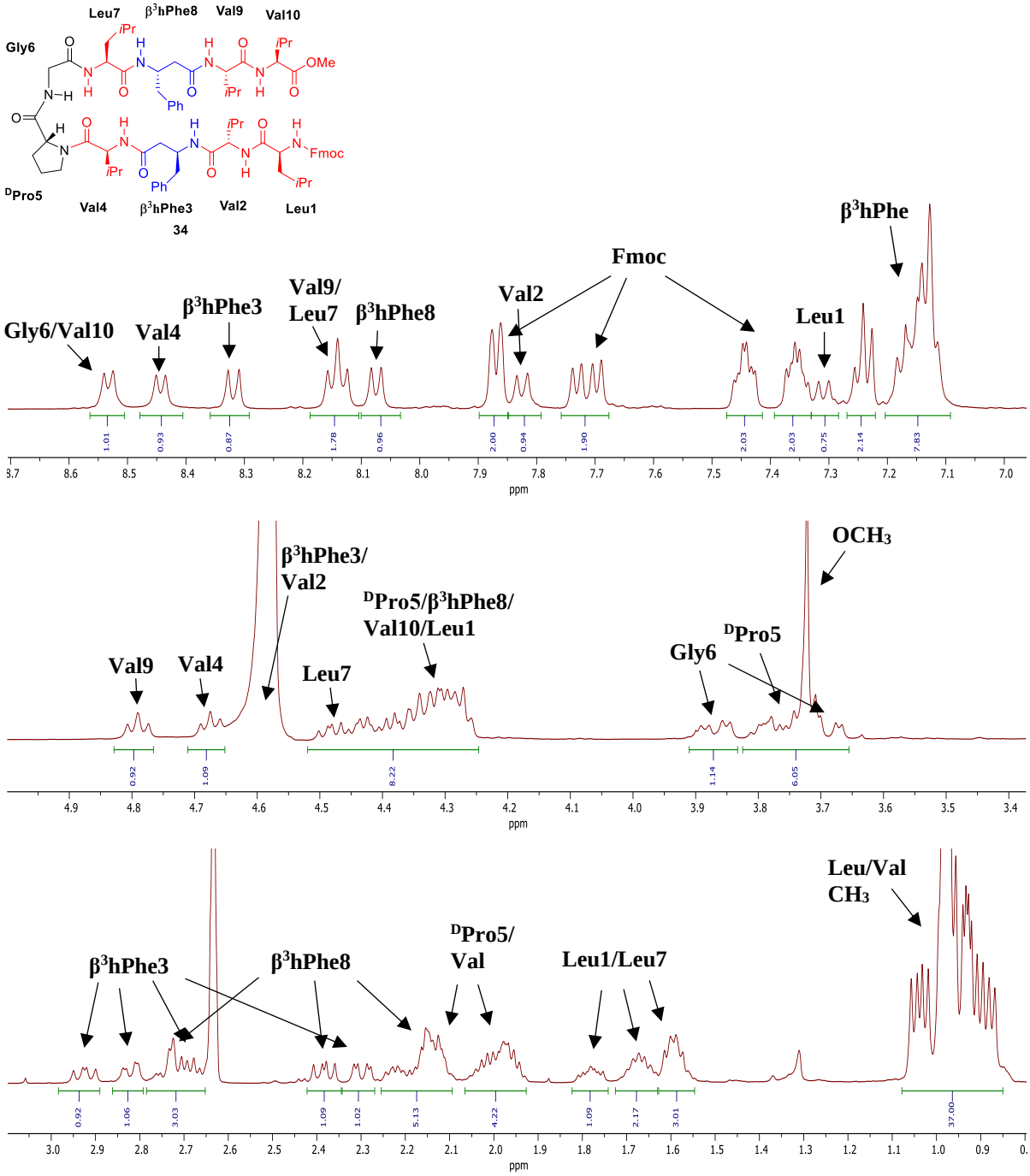


Figure 3.4: 500 MHz ^{1}H NMR spectrum of peptide **34** (~4 mM) in $CD_3OH/(CD_3)_2SO$ (3:1). **Top**. NH and aromatic CH (Fmoc and β^3hPhe) region. **Middle**. Amino acid $C^{\alpha}H$ region. **Bottom.** β^3hPhe benzylic CH_2- and $C^{\alpha}H_2$, aliphatic CH_3, CH_2 and CH (Val, Leu, ^{D}Pro) region.

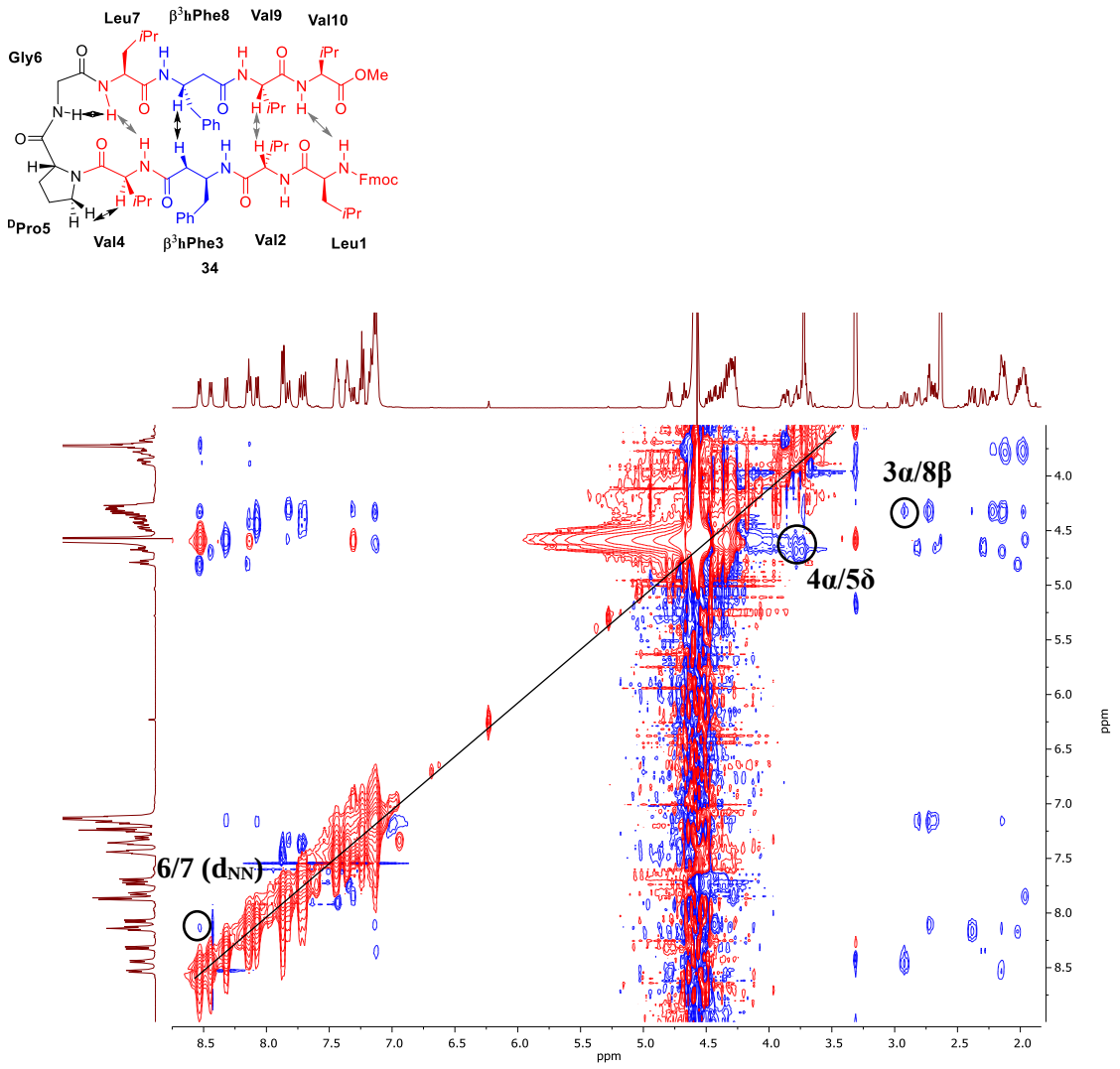


Figure 3.5: Partial ROESY NMR spectrum of peptide **34** in $CD_3OH/(CD_3)_2SO$ (3:1). Relevant NOEs are appropriately labelled. Embedded is a schematic of **34**, double-edge arrows in black indicate the observed sequential and cross-strand NOEs, double-edge arrows in grey indicate missing cross-strand NOEs observed for peptide **5**.

On the other hand, having synthesised mismatched peptide **36**, we attempted to evaluate its secondary structure in solution. To do so, the ^{1}H NMR spectrum of **36** in CD_3OH was unambiguously assigned using the same NMR experiments carried out on matched counterpart **34**. The mismatched peptide was less soluble in CD_3OH than peptide **34**, saturating at ~1 mM. However, given the possibility that $(CD_3)_2SO$ could disfavour the formation of intramolecular β-sheets, it was decided to perform the conformational studies at this concentration without a co-solvent.

A fully ^{1}H NMR spectrum of mismatch **36** is shown in Figure 3.6. As it was the case for match **34**, the NH resonances are well dispersed (ranging from δ = 7.0 to 8.6 ppm). In the $C^{\alpha}H$ region, signals are dispersed over the range of δ = 4.1 to 5.3 ppm. Unlike **34**, the HSQC NMR spectrum revealed none of the backbone $C^{\alpha}H$ signals overlapped with the residual water signal.

According to the literature, peptides with random conformations normally present backbone NH groups with $^{3}J_{NH\text{-}C\alpha H}$ values of around δ = 7 Hz.[93] Since we hypothesise that peptide **36** will present such an unordered secondary structure it was unexpected to observe that all the NH groups present $^{3}J_{NH\text{-}C\alpha H}$ values of δ ≥ 8 Hz, closer to what is expected for ordered β-sheet conformations.

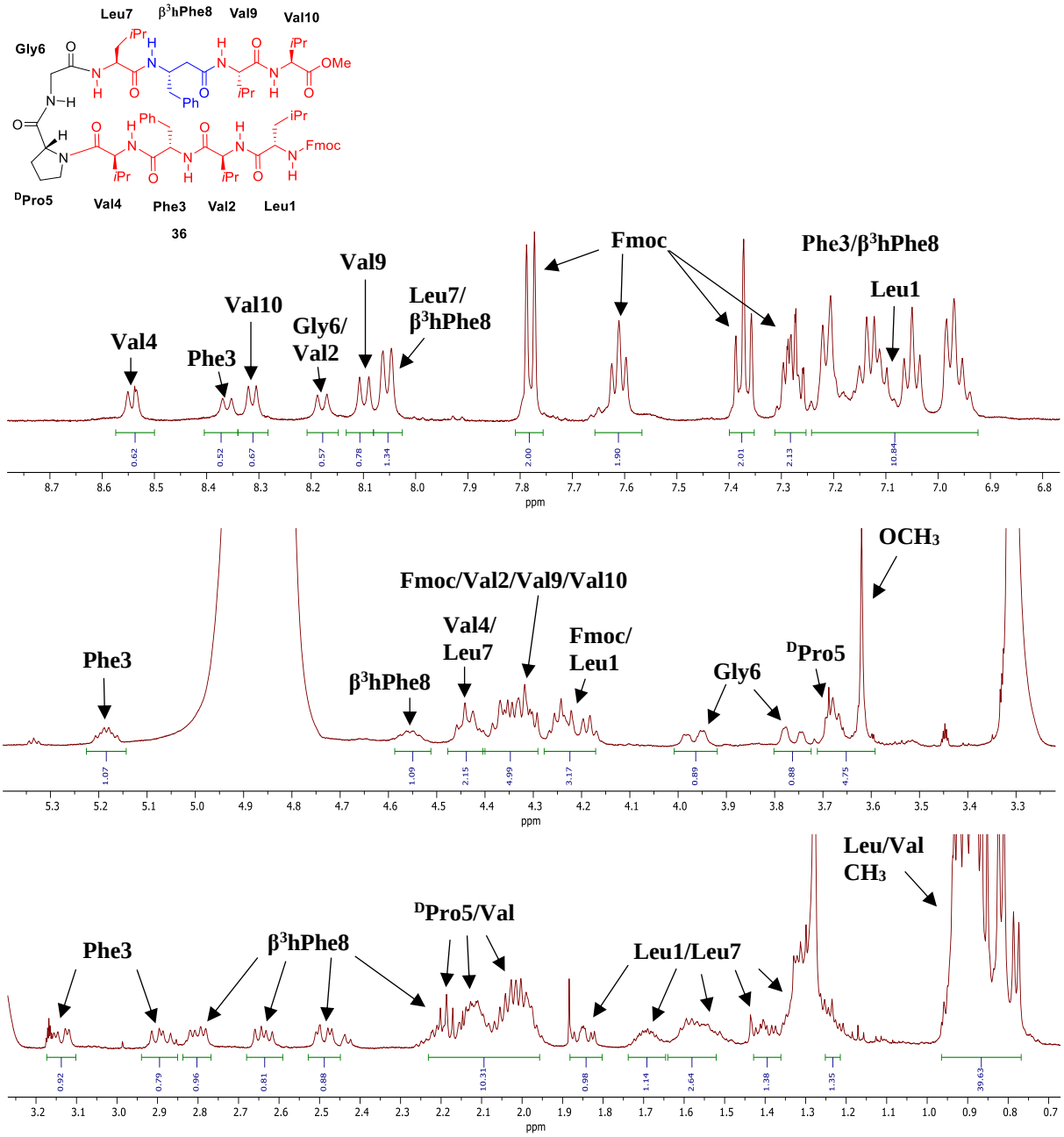


Figure 3.6: 500 MHz ^{1}H NMR spectrum of peptide **36** (~1 mM) in CD_3OH. **Top**. NH and aromatic CH (Fmoc, Phe3, β^{3}hPhe8) region. **Middle**. Amino acid C$^{\alpha}$H region. **Bottom**. Phe3, β^{3}hPhe8 benzylic CH_2 and C$^{\alpha}H_2$, aliphatic CH_3, CH_2 and -CH (Val, Leu, ^{D}Pro) region.

To assess the peptide's conformation in solution a ROESY NMR spectrum of mismatch **36** was recorded using the same spectroscopic parameters used for match **34**. Figure 3.7 shows partial a ROESY

spectra with highlighted sequential backbone NOE interactions observed for **36**. Unlike **34**, no cross-strand contacts were observed for the mismatched peptide. Figure 3.7b, shows a ^{D}Pro5 C$^{\delta}$H↔Val4 C$^{\alpha}$H sequential contact which could be correlated with a β-turn conformation. However, there is no evidence for the presence of diagnostic Gly6↔Leu7 (d_{NN}, i + 2/i + 3) contact to confirm this structural feature.

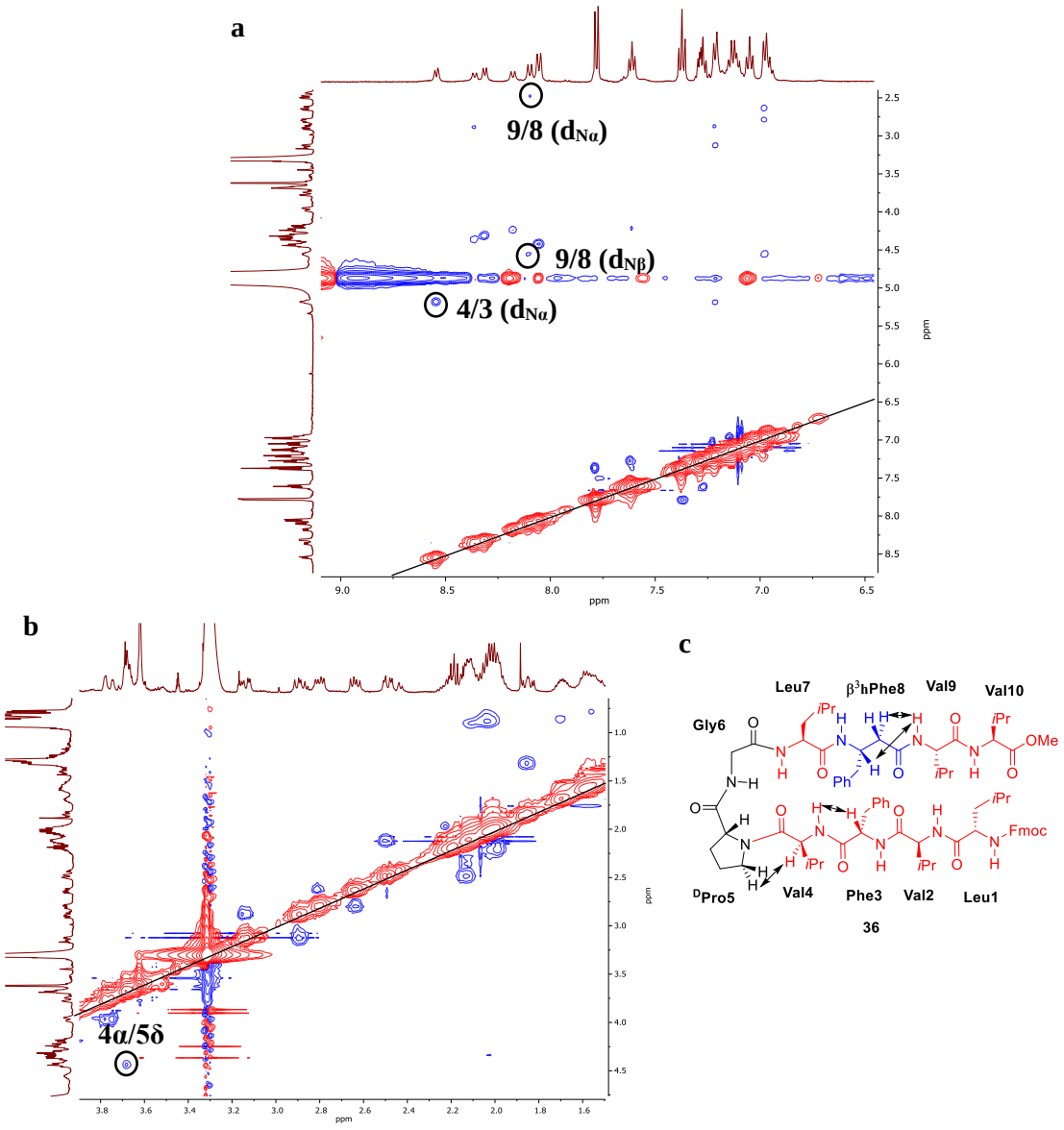
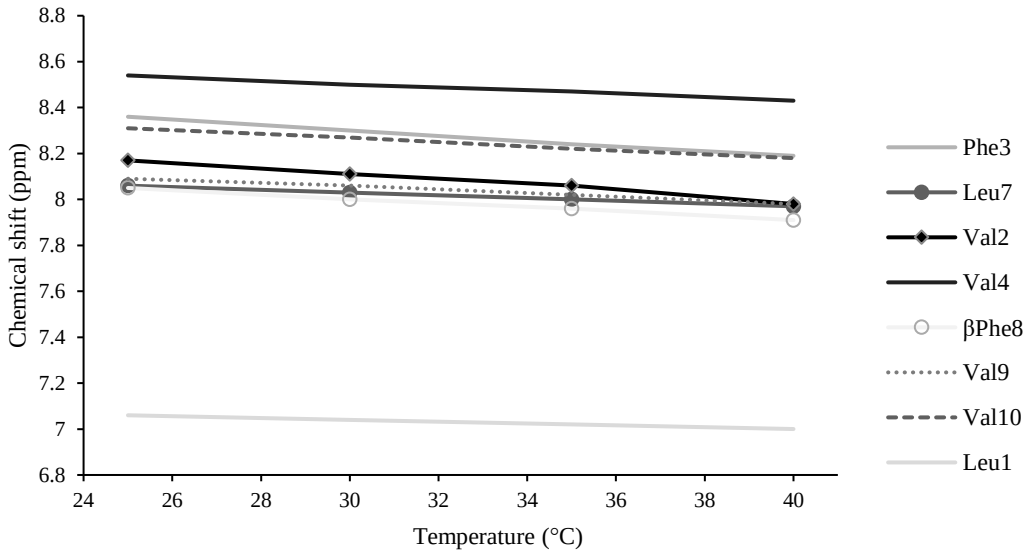


Figure 3.7: **a**. Partial ROESY NMR spectra of peptide **36** in CD_3OH, relevant sequential backbone NOEs are appropriately labelled. **b**. Partial ROESY NMR spectra highlighting a 4Val↔^{D}Pro5 NOE **c**. A schematic of **36**, double-edge arrows indicate the observed sequential NOEs.

Another NMR-based study used to evaluate intramolecular hydrogen-bonding is the measurement of the temperature dependence of amide chemical shifts. There is literature precedent to the idea that the chemical shifts of amide protons involved in intramolecular hydrogen-bonding are less influenced by a change in temperature than amide protons that are hydrogen-bonded only with solvent.[95] Thus, to gain further insight on the conformation adopted by peptide **36**, the temperature dependence of NH chemical

shift in CD_3OH was measured (see section 7.2.7), spectra recorded at the different temperatures are shown in section 8.4). All amides exhibited linear temperature dependence (Figure 3.8, top). The dδ/dT values (ppb/°C) are listed in the table on Figure 3.8. The NH resonances of virtually all the residues exhibit high dδ/dT values of > 5 ppb/°C, characteristic of solvent-exposed NH resonance, which suggests the residues are not involved in intramolecular hydrogen bonds. The NH group of the Gly6 residue could not be followed due to an absence of the $NH \leftrightarrow C^{\alpha}H_2$ coupling in the TOCSY NMR spectrum at high temperatures.



Residue	dδ/dT (ppb/°C)
Phe3	11.4
Leu7	6
Leu1	4
Val2	12.4
Val4	7.2
β^3hPhe8	9.2
Val9	7.4
Val10	8.8

Figure 3.8: **Top**. Temperature dependence of NH chemical shifts of peptide **36** in CD_3OH. **Bottom**. Table of NH temperature coefficients for peptide **36** in CD_3OH.

Overall, NOE data so far pointed to the matched peptide **34** folding into a β-hairpin conformation, but there was little evidence of mismatched peptide **36** adopting such conformation. However, we realised that it would be difficult to obtain more data (especially NOE patterns) using CD_3OH, and that $(CD_3)_2SO$ was not viable, so instead we set out to evaluate the conformation of the α/β-peptides in a different solvent.

3.3.4. Conformational studies of matched decamers 34 and 35 by NMR in $CDCl_3$

$CDCl_3$ was chosen as an alternative solvent to methanol since it had been used before by Balaram in the characterization of similar peptide β-hairpins.[92] Moreover, this solvent should not interfere with intramolecular hydrogen-bonding and, like CD_3OH, it would allow us to monitor the NH↔NH NOEs. Due to our experimental design, it made sense to characterise the matched sequences in the new solvent, look for the spectroscopic patterns reported for β-hairpin motifs and then compare them with the mismatched sequences. We began by evaluating match **35** in $CDCl_3$.

Pleasingly, peptide **35** was found to be very soluble in $CDCl_3$. A ~4 mM solution was prepared and a ^{1}H NMR spectrum was recorded. The fully assigned proton spectrum is shown in Figure 3.9. a very well dispersed group of NH resonances (ranging from δ = 5.8 to 8.8 ppm) was observed, although with some degree of overlap, especially on the signals corresponding to Val8, Val3 and β^3hPhe2. The $C^{\alpha}H$ region, also showed good dispersion (δ = 4.1 to 5.0 ppm).

Unlike the spectrum of matched decamer **34** in CD_3OH, at this concentration of match **35**, $CDCl_3$ caused peak broadening in both backbone NH and $C^{\alpha}H$ groups, this led to a loss of splitting patterns and consequently, no information about the $^3J_{NH\text{-}C\alpha H}$ values could be extracted from the 1D ^{1}H-NMR spectrum. Nevertheless, given the good dispersion of the signals it was decided to study the NOE patterns displayed by **35** in this solvent.

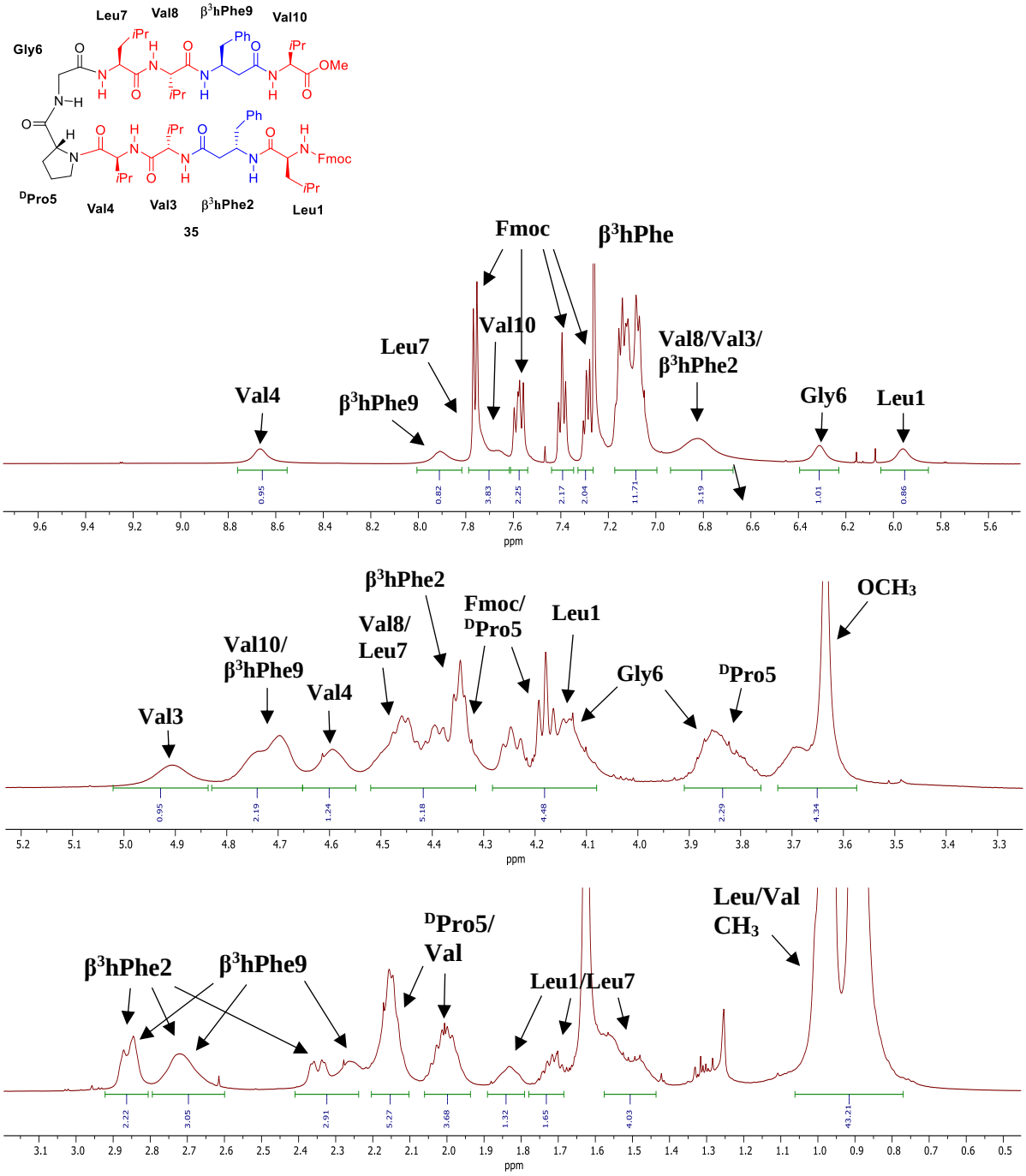


Figure 3.9: 500 MHz ^{1}H NMR spectrum of peptide **35** (~4 mM) in $CDCl_3$. **Top**. NH and aromatic CH (Fmoc, β^{3}hPhe) region. **Middle**. Amino acid C$^{\alpha}$H region. **Bottom.** β^{3}hPhe benzylic CH_2 and C$^{\alpha}H_2$, aliphatic CH_3, CH_2 and CH (Val, Leu, ^{D}Pro) region.

ROESY and NOESY NMR spectra of match **35** were recorded using the same spectroscopic parameters used for match **34**. Both experiments showed roughly the same contacts. The partial NOESY in Figure 3.10a shows a series of cross-strand NOEs expected in an antiparallel sheet conformation; Val3 C$^{\alpha}$H↔Val8 C$^{\alpha}$H, β^{3}hPhe2 C$^{\beta}$H↔β^{3}hPhe9 C$^{\alpha}$H, Leu1↔Val10 ($d_{N\beta}$) and a contact between the methoxy group in Val10 and an aromatic proton in the Fmoc group of Leu1. Figure 3.10b shows partial NOESY

spectra where the sequential NOE Gly6↔Leu7 (d_{NN}) characteristic for type II' β-turn at the ^{D}Pro-Gly segment is observed. The figure also shows a cross-strand Val4↔Leu7 (d_{NN}) contact. These NOE interactions provide strong evidence of peptide **35** folding into a hairpin in $CDCl_3$.

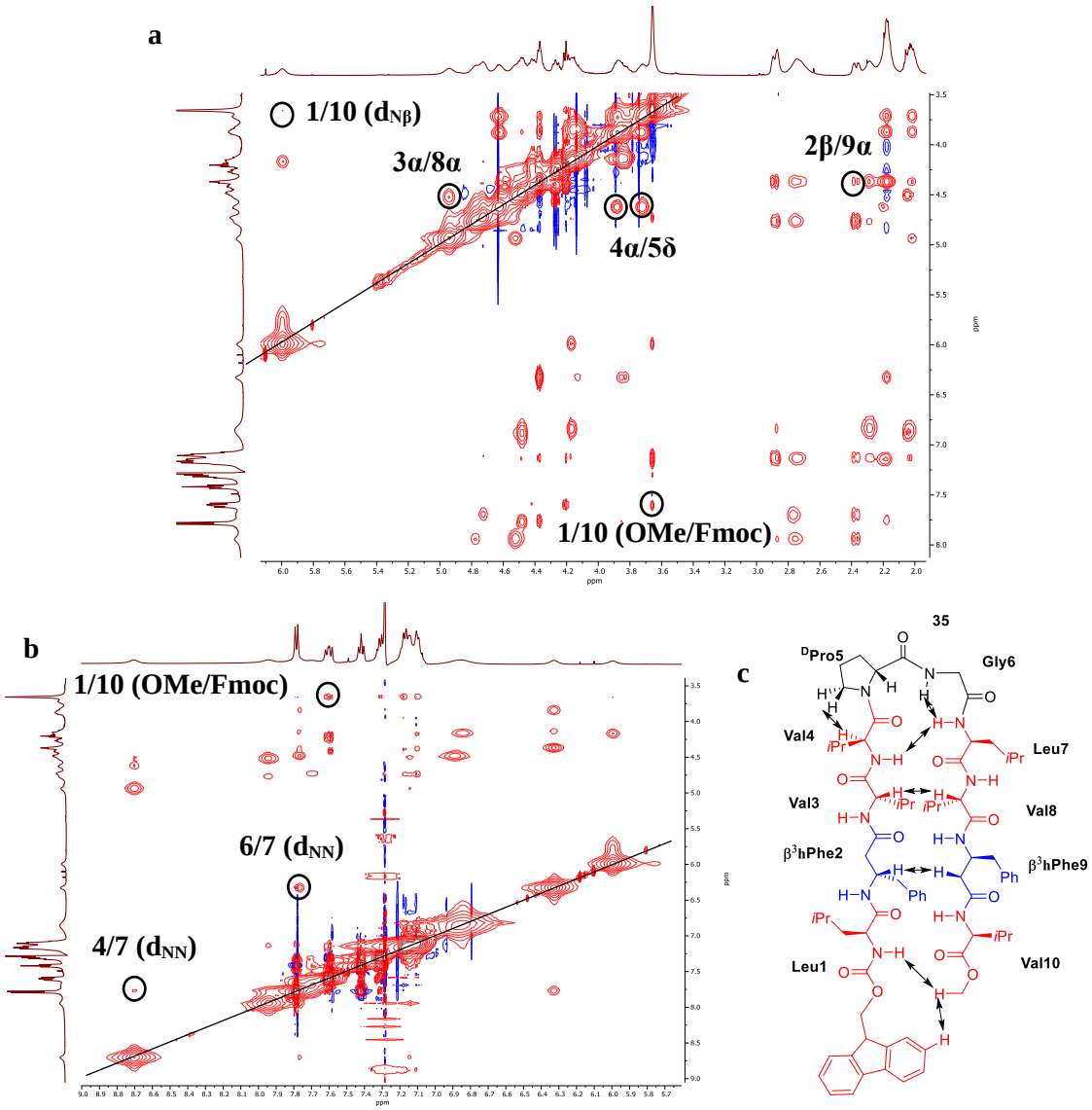
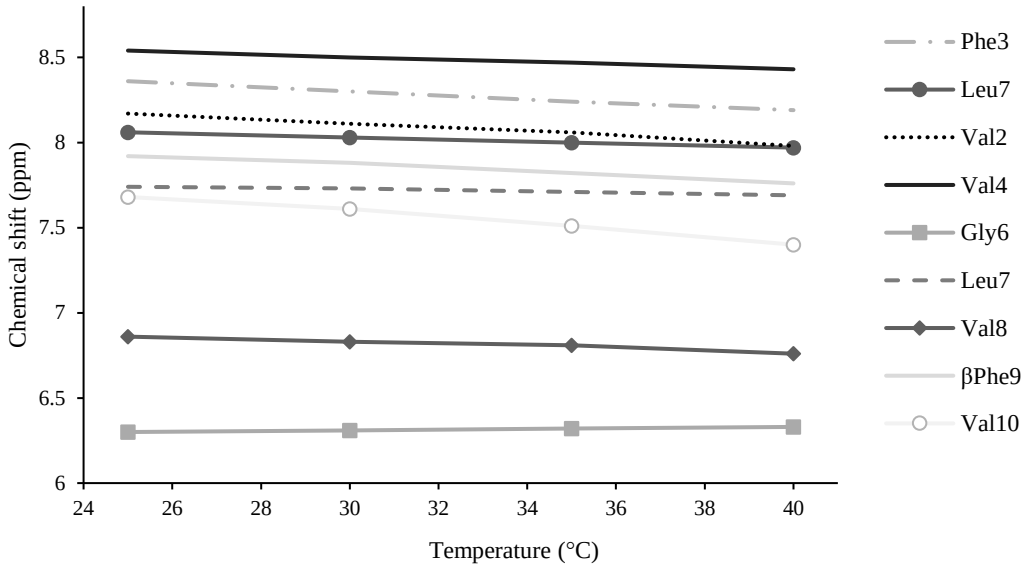


Figure 3.10: **a** Partial NOESY NMR spectra of peptide **35** in $CDCl_3$, relevant cross-strand NOEs are appropriately labelled. **b**. Partial NOESY NMR spectra highlighting relevant NOEs. **c**. A schematic of **35**, double-edge arrows indicate the observed sequential and cross-strand NOEs.

In addition to the NOE experiments, the temperature dependence of chemical shifts was measured for match **35** (Figure 3.11, spectra recorded at the different temperatures are shown in section 8.4). The experiment showed that of the expected internally hydrogen-bonded NH groups in this conformation, only Leu7 (residue *i* + 3 of the β-turn) presented low temperature coefficients (Figure 3.11). However, Leu1, Val4, β^3hPhe9 and Val10 displayed high temperature coefficients, unexpected for intramolecular hydrogen-bonded NH groups. These observations were ambiguous for a β-hairpin conformation. We rationalised that our study was probably not covering a large enough temperature range (literature

studies cover ranges of around 0-40 °C)[92] and that could be causing the inconsistent pattern. Moreover, it has been shown that for many peptides, the temperature coefficient could mainly describe the temperature-dependent loss of folded structure and only indirectly characterise hydrogen-bonding.[95]



Residue	dδ/dT (ppb/°C)
Leu1	11.4
β^3hPhe2	4.8
Val3	9
Val4	8
Gly6	2
Leu7	3.4
Val8	6.8
β^3hPhe9	10.8
Val10	18.8

Figure 3.11: **Top**. Temperature dependence of NH chemical shifts of peptide **35** in $CDCl_3$. **Bottom**. Table of NH temperature coefficients for peptide **35** in $CDCl_3$.

While studying match **35**, we became interested in obtaining NMR spectra at lower concentrations in $CDCl_3$ to study the molecule's conformation. We anticipated that diluted samples might result in better resolved NOEs and a more informative 1D ^1H NMR spectrum. Furthermore, dilution studies have been conducted in the literature for the characterization of peptide hairpin analogues.[92]

While recording ^1H NMR and NOESY NMR spectra of **35** at different concentrations, change in NH chemical shifts became apparent, which hinted at the possibility of match **35** forming extended pleated-sheet aggregates at high concentrations in $CDCl_3$. Under this conditions, β-hairpins bearing matching α/β-sequences would be expected to form infinite head to head aggregates (Figure 3.12).

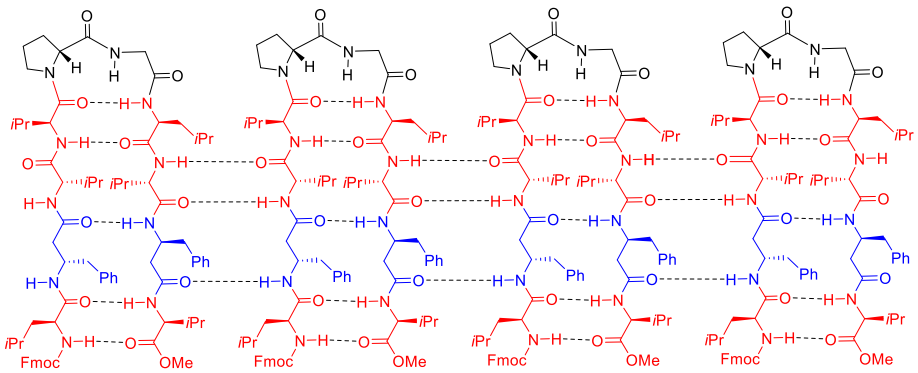
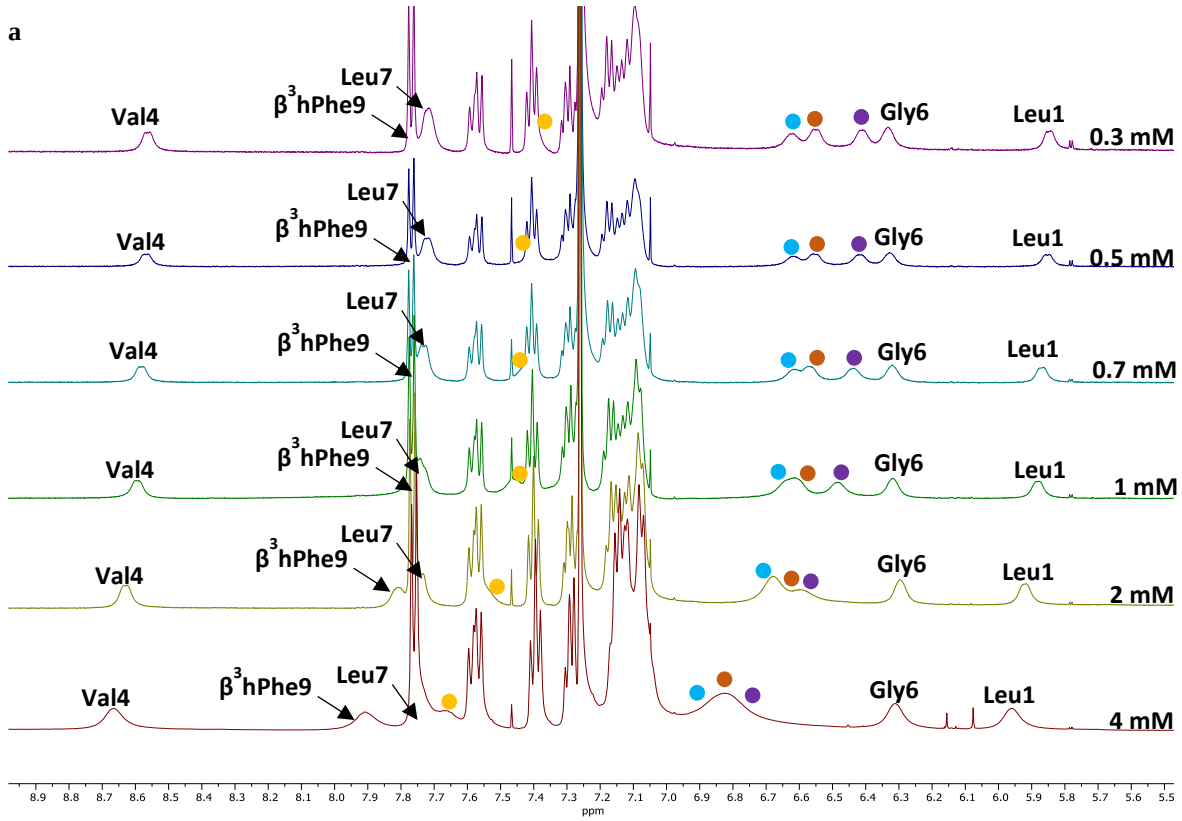


Figure 3.12: Assembly of peptide **35** into an extended β-sheet by intermolecular hydrogen-bonding.

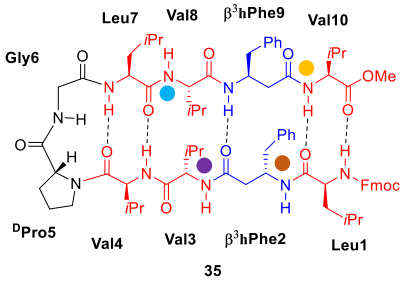
To explore this hypothesis, a sequential dilution study was carried out. ^1H NMR spectra were recorded at six different concentrations, ranging from 4 to 0.3 mM (Figure 3.13a). Upon monitoring the effect of the concentration on the ^1H chemical shift, it was noticed that most of the δ_{CH} remained virtually unchanged throughout the dilution. However, some amide protons did present a variation in their chemical shifts as the concentration changed; as shown in Figure 3.13b, the effect was not the same for all the NH groups. Interestingly, amide groups from residues Val4, Leu7 and Leu1 anticipated to participate in intramolecular hydrogen-bonding in the hairpin conformation presented the smallest Δδ (0.01-0.17 ppm), this was particularly true for those nearest to the β-turn (Figure 3.13c). Conversely, the largest changes to the NH chemical shift were exhibited by two of the amides at non-hydrogen-bonding positions in the β-hairpin conformation: Val3 and Val8 ($\Delta\delta_{NH}$ = 0.36 and 0.30, respectively; Figure 3.13c). It was also noted that as the sample concentration decreased, the NH signals sharpened, and splitting started to become apparent.

These results were interpreted as indicative of matched decamer **35** forming intermolecular assemblies at high concentrations. As mentioned in section 3.1, peptides folding into β-hairpins have been shown to form extended sheets under certain conditions.[75,76] Thus, our observations can be rationalised by considering that at higher concentrations, lateral aggregation of the peptide β-strands through solvent-exposed amide groups to form extended sheets will occur through hydrogen-bonding. As the concentration of **35** decreases, these intermolecular associations are disfavoured, and this is reflected in

a change on the δ_{NH} of solvent-exposed amide groups. The occurrence of this aggregation process would strengthen the evidence for match **35** adopting a folded β-hairpin conformation and at the same time, for PMC information transfer.



b



35

c

Residue	$\Delta\delta_{NH}$ (ppm)
Leu1	0.11
β³hPhe2	0.19
Val3	0.36
Val4	0.10
Gly6	-0.02
Leu7	-0.01
Val8	0.30
β³hPhe9	0.17
Val10	0.24

Figure 3.13: **a**. Partial ^{1}H NMR spectrum of peptide **35** in $CDCl_3$ at four different concentrations, the changes in δ_{NH} are highlighted. The NH of Val10 residue is denoted with ● , the NH of Val8 residue is denoted with ● , the NH of Val3 residue is denoted with ● and the NH of β³hPhe2 residue is denoted with ● . **b**. Structure of **35** illustrating internal hydrogen bonds. **c**. Table summarizing the $\Delta\delta_{NH}$ observed in the dilution study.

After these positive results using $CDCl_3$ and having matched decamer **34**, its conformation in chloroform was studied by NMR. As it was the case for match **35**, peptide **34** was found to very soluble in $CDCl_3$. A ~4 mM solution was prepared and a ^1H NMR spectrum was recorded. The fully assigned proton spectrum is shown in Figure 3.14. A very well dispersed group of NH resonances (ranging from δ = 6.3 to 8.6 ppm) with some degree of overlap, especially on the signals corresponding to Val9, Val2 and β^3hPhe8 was observed. As it had been noticed with peptide **35**, peak broadening in both backbone NH and $C^{\alpha}H$ groups caused a loss of splitting patterns and consequently, no information about the $^3J_{NH\text{-}C\alpha H}$ values could be extracted from the 1D ^1H-NMR spectrum.

When comparing the spectra obtained for match **34** in the two solvents (CD_3OH and $CDCl_3$), it was noticed that the chemical shift of the NH signals corresponding to residues Val9, Val2 and β^3hPhe8 presented the largest variations when going from CD_3OH to $CDCl_3$ (roughly -1ppm). This was expected, given that these amides should be solvent-exposed in the hairpin conformation, and therefore would appear at lower fields due to hydrogen-bonding interactions with CD_3OH. On the other hand, it was observed that the chemical shift of the NH signals corresponding to residues Val10, Val4, Leu7 and β^3hPhe3 remained practically unchanged, this correlates well considering that these amides are expected to be engaged in intramolecular hydrogen-bonding in the hairpin conformation. The amide from the Leu1 residue presented a large change in chemical shift when going from CD_3OH to $CDCl_3$ (-0.9 ppm), hinting at some fraying of the intramolecular β-sheet at the most distant point of the β-turn.

ROESY and NOESY NMR spectra of match **34** were then recorded using the same spectroscopic parameters used for match **35**. Both experiments showed roughly the same contacts. The partial NOESY in Figure 3.15a-b shows a series of cross-strand NOEs expected in an antiparallel sheet conformation; Val2 $C^{\alpha}H$↔Val9 $C^{\alpha}H$, β^3hPhe3 $C^{\alpha}H$↔β^3hPhe8 $C^{\beta}H$, Leu1↔Val10 ($d_{N\beta}$). The figure also shows the sequential NOEs ^{D}Pro5 $C^{\delta}H$↔Val4 $C^{\alpha}H$ and Gly6↔Leu7 (d_{NN}) characteristic for type II' β-turn at the ^{D}Pro-Gly segment. These NOE interactions provide strong evidence of peptide **34** folding into a hairpin in $CDCl_3$.

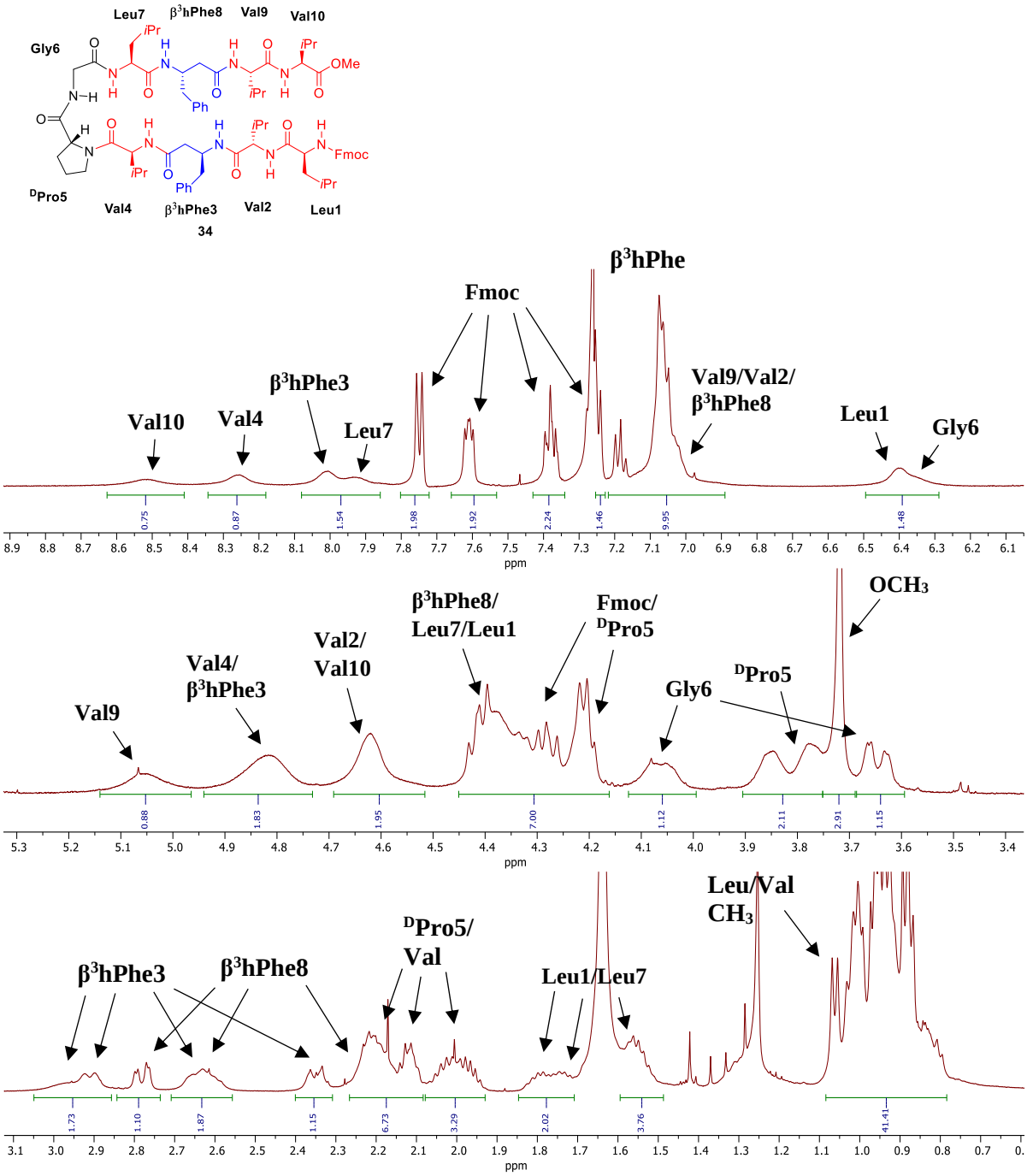


Figure 3.14: 500 MHz ^{1}H NMR spectrum of match **34** (~4 mM) in $CDCl_3$. **Top**. NH and aromatic CH (Fmoc, β^{3}hPhe) region. **Middle**. Amino acid C$^{\alpha}$H region. **Bottom.** β^{3}hPhe benzylic CH_2 and C$^{\alpha}H_2$, aliphatic CH_3, CH_2 and CH (Val, Leu, ^{D}Pro) region.

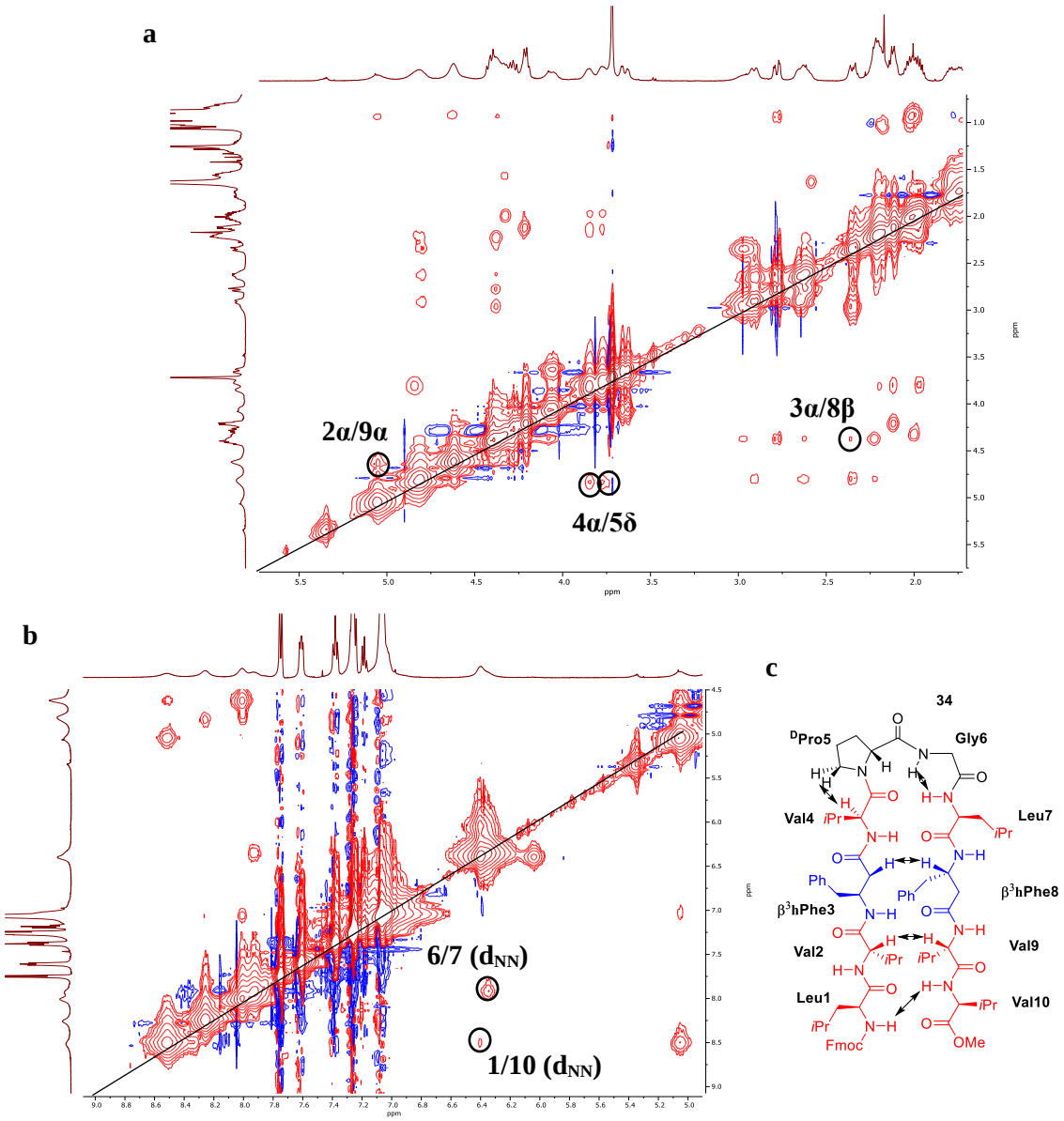


Figure 3.15: **a**. Partial NOESY NMR spectra of peptide **34** in $CDCl_3$, relevant cross-strand NOEs are appropriately labelled. **b**. Partial NOESY NMR spectra highlighting relevant d_{NN} NOEs. **c**. A schematic of **34**, double-edge arrows indicate the observed sequential and cross-strand NOEs.

Given that the study of the temperature dependence of the amide chemical shift had not provided robust data towards the conformation adopted by the peptide, it was omitted for peptide **34**. Instead, with the information that had been obtained for the two matches we focused on studying the conformation of the mismatched peptides.

3.3.5. Difference in the solubility of mismatched decamers 36 and 37

We were interested in studying the secondary structure of the mismatched peptide sequences in $CDCl_3$. However, mismatched peptide **36** presented limited solubility in this solvent; formation of visible aggregates was noticed upon addition of chloroform (Figure 3.16a). Surprisingly, after short periods of time, the suspension also acquired a yellowish colour (Figure 3.16b). Upon examining the ^1H NMR spectrum of this dilute sample it was observed that more than one peptide species was present in solution. Cleavage of the Fmoc-protecting group at the N-terminus was detected. Although this prevented us from obtaining more conclusive spectroscopic data, we rationalised that these observations supported the hypothesis that mismatch **36** was adopting an alternative (potentially less ordered) conformation to the β-hairpin motif, since it is generally regarded that a high solubility in apolar solvents such as chloroform is an indication of the formation of folded structures, stabilised by intramolecular hydrogen bonds.[92]

On the other hand, the behaviour of mismatched peptide **37** was remarkably different to that of matched isomers **34** and **35**. After the SPPS, the peptide could not be identified by ESI or MALDI-TOF MS, and it was only detected in the crude mixture by NSI MS at $[M+H]^+$ = 1353.7 *m/z* (calculated $[M+H]^+$ = 1353.8 *m/z*). However, the peptide could not be isolated due to its virtual insolubility in most common solvents. A wide variety of additives were used in an attempt to obtain an analytical HPLC trace of the crude peptide. This included DMF, DMSO, 0.1% TFA, GndmCl, TFE, as well as, MOPS and NH_4CO_3 buffers at basic pHs. Our attempts were unsuccessful though, and further characterization could not be carried out.

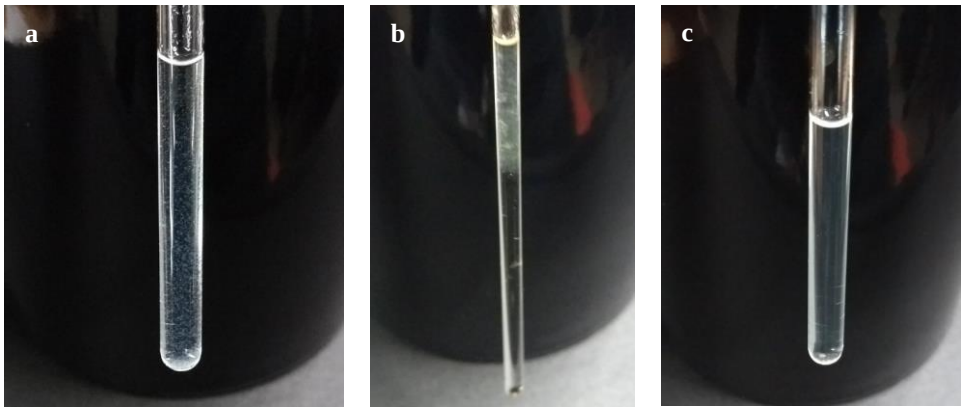


Figure 3.16: **a**. Suspension of peptide **36** in $CDCl_3$. **b**. Suspension of peptide **36** in $CDCl_3$ after 10 minutes of solvent addition. **c**. A 4 mM solution of peptide **35** in $CDCl_3$ shown for comparison.

3.3.6. An alternative mismatched decamer

Failure to study mismatch **37** prompted us to design an alternative mismatched variant. Thus, we synthesised decamer **38** (Figure 3.17), an isomer of mismatch **37** and matches **34** and **35** where the β-residues are only one residue away from each other in the facing strands. This resulted in merely one pair of unbound H-bonding partners in the middle of the hairpin strands.

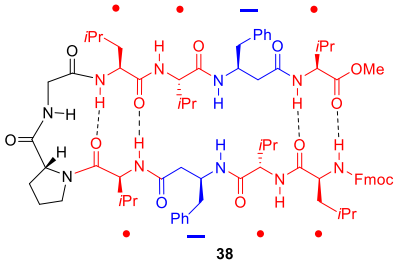
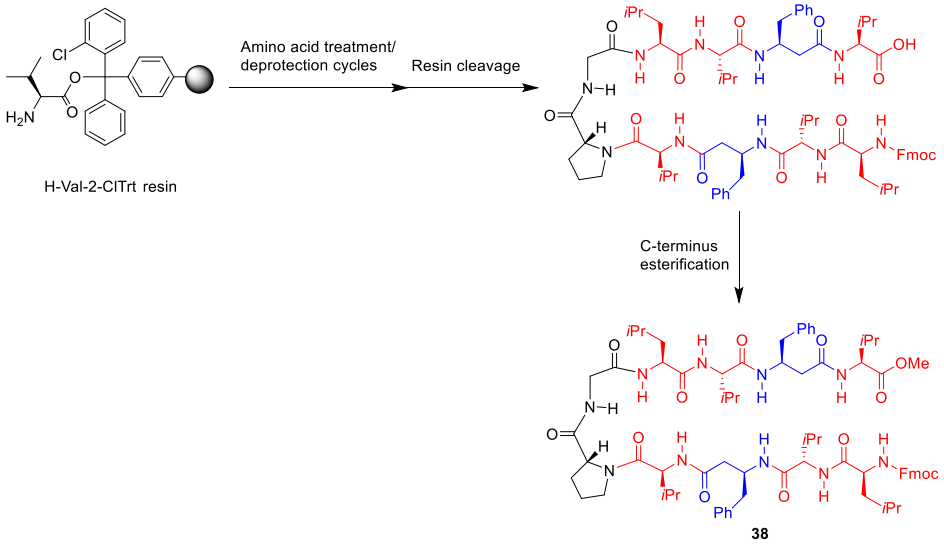


Figure 3.17: Design for decapeptide **38** bearing mismatching α/β-strands.

The peptide was synthesised using a standard manual Fmoc-based SPPS methodology to yield the C-terminal free carboxylic acid. This was then converted into a methyl ester derivative using the same methodology employed with the previous decamers (Scheme 3.3). The peptide was purified through preparative HPLC (methanol/water). The identity and purity of the peptide was assessed by MS and analytical HPLC.



Scheme 3.3: Overview of the synthesis of decamer **38** (for full conditions see sections 7.2.1 and 7.3.2).

Having synthesised mismatch **38**, we proceeded to record a ^{1}H NMR spectrum in $CDCl_3$. It was noted that the peptide was not as soluble as its matched counterparts, but a ~2 mM solution in this solvent could still be obtained. This solution yielded the ^{1}H NMR spectrum shown in Figure 3.18. The spectrum showed great similarity to that of matched decamers **34** and **35**, the major differences were displayed

by some of the amide groups; specially those corresponding to the Val4 and Val10 residues, which appear at higher fields relative to the matched hairpin isomers (Figure 3.18). This is an interesting observation, considering those amides are expected to be intramolecularly H-bonded in a β-hairpin conformation.

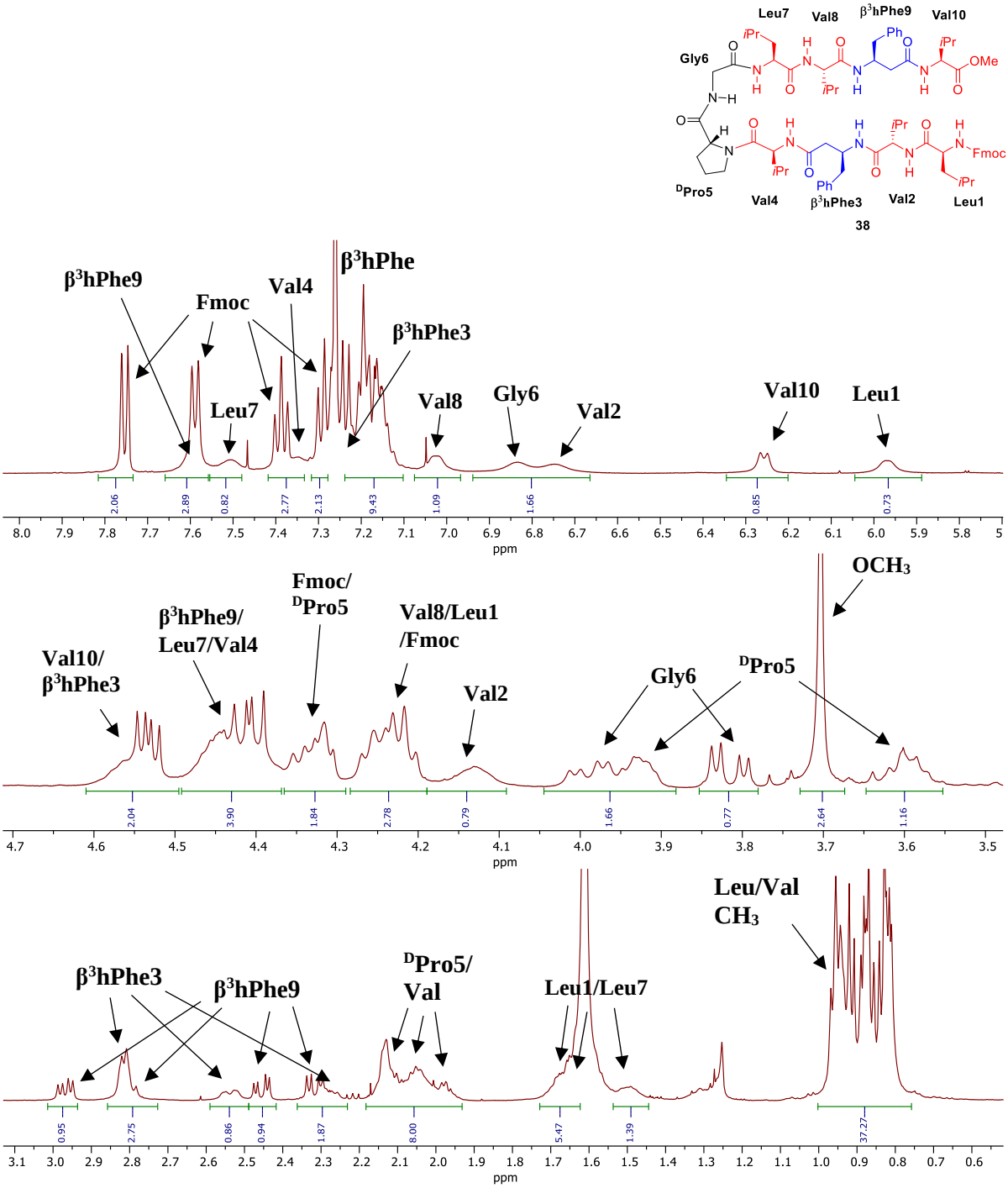


Figure 3.18: 500 MHz ^{1}H NMR spectrum of peptide **38** (~2 mM) in $CDCl_3$. **Top**. NH and aromatic CH (Fmoc, β^{3}hPhe) region. **Middle**. Amino acid C$^{\alpha}$H region. **Bottom.** β^{3}hPhe benzylic CH_2 and C$^{\alpha}H_2$, aliphatic CH_3, CH_2 and CH (Val, Leu, ^{D}Pro) region.

ROESY and NOESY NMR spectra of mismatch **38** were then recorded using the same spectroscopic parameters used for matches **34** and **35**. The ROESY spectra showed a better pattern of correlations than NOESY. The partial ROESY spectra in Figure 3.19 show a series of sequential backbone NOE interactions observed for **38**. Figure 3.19a, shows a ^{D}Pro5 C$^{\delta}$H↔Val4 C$^{\alpha}$H sequential contact and Figure 3.19b shows a weak Gly6↔Leu7 (d_{NN}) contact, both correlating to a β-turn conformation. However, unlike it was the case for **34** and **35**, none of the cross-strand contacts expected for a folded β-sheet-like conformation were observed for mismatch **38**. Besides, the weak Val10 NH↔ β^{3}hPhe9 C$^{\beta}$H suggests a less rigid folded conformation near the β-residue at the strand terminus. Notably, an analogous (Val9 NH↔ β^{3}hPhe8 C$^{\beta}$H, *vide supra*) but more intense NOE was also observed in mismatch **36**.

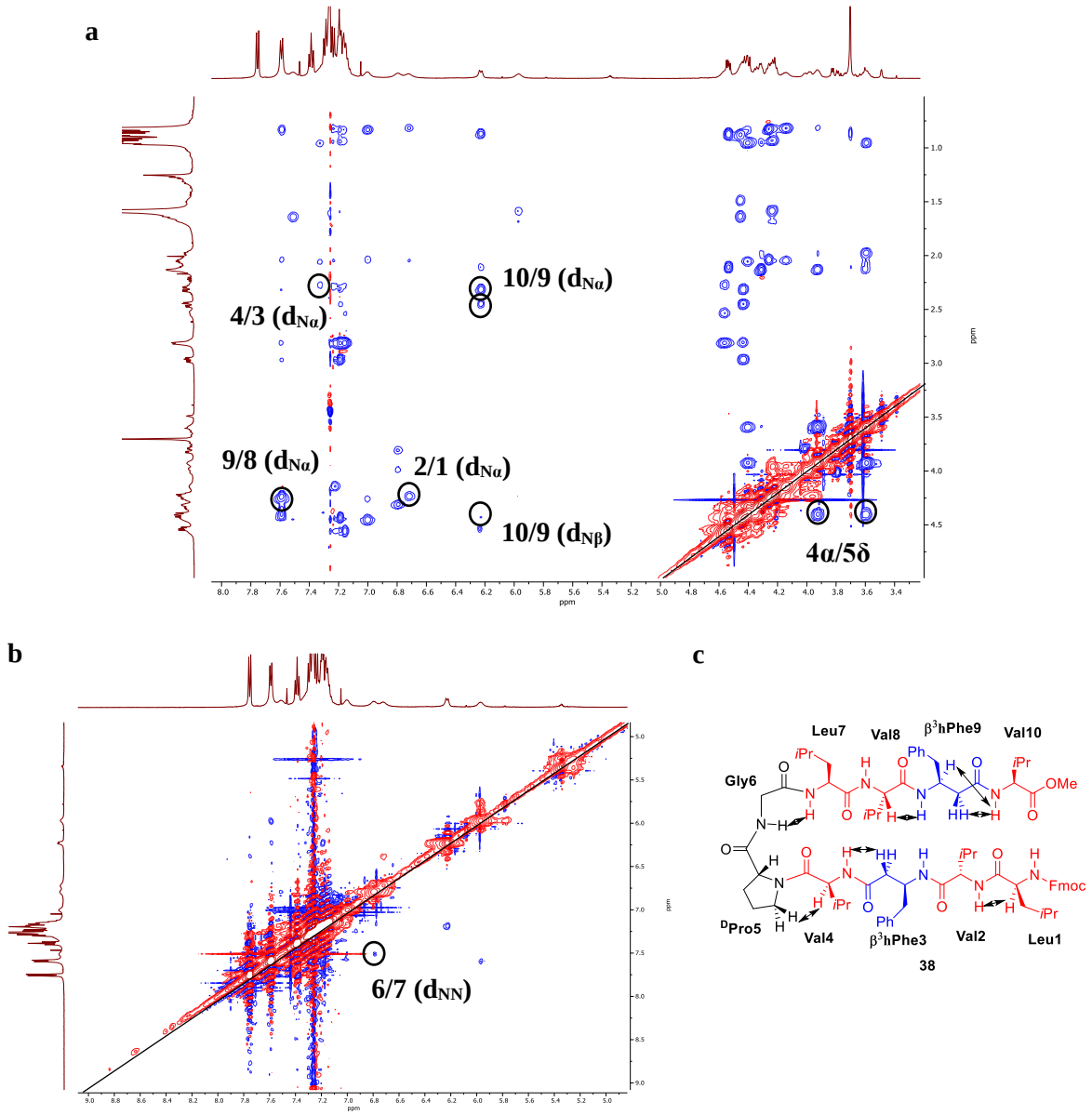


Figure 3.19: **a** Partial ROESY NMR spectra of peptide **38** in $CDCl_3$, relevant sequential backbone NOEs are appropriately labelled. **b**. Partial ROESY NMR spectra highlighting the Gly6↔Leu7 (d_{NN}) NOE **c**. A schematic of **38**, double-edge arrows indicate the observed sequential NOEs.

Overall, NOE analyses pointed to matched decamers **34** and **35** folding into β-hairpins in chloroform, whereas mismatching isomer **38** presented less ordered structures, caused by the fraying of the strands due to inefficient intramolecular hydrogen-bonding.

3.3.7. Conformational studies of decamer α/β-peptide β-hairpins by circular dichroism

The folding behaviour of the decamer α/β-peptides **34**-**38** was also monitored by circular dichroism, a technique commonly used to characterise the secondary structure of short peptides. Figure 3.20 shows the CD spectra of the four peptides in the series at concentrations close to 100 μM using methanol as a solvent (due to the incompatibility of chloroform with far-UV CD, the conformation adopted by our molecules could not be monitored in this solvent).

Match **34**, which is based directly on literature compound **5** (characterised as a type II' β-turn nucleated hairpin in solution), displayed virtually the same spectra as the literature compound, with a minimum at approximately 220 nm and a maximum at 198 nm, a CD signature of peptide β-hairpins.[75,96] The spectrum of isomeric match **35** showed the same pattern of bands, but with different intensities. The negative band close to 220 nm has been used as an indicator to estimate the content of β-sheet secondary structure;[97,98] comparison of the intensity of this band suggests that the residue-normalised population of β-sheet structure is greater in **35** relative to **34**. This was an interesting observation, since it was expected that match **34**, bearing the flexible β^3hPhe residues closer to the turn segment, would generate the most stable sheet conformation. This was interpreted as an indication that the position of the matching β-residues was not affecting the stability of the secondary structure in these short peptides.

In the case of mismatch **36**, the CD spectrum showed a very different pattern to that of the two matches. A negative intensity but no defined minimum is observed in the 220 nm region, which is consistent with a small content of folded structure, and an intense negative band is observed around 200 nm, a characteristic band for a peptide predominantly adopting a statistically unordered conformation.[94]

The CD spectrum of mismatch **38** is less conventional than those displayed by peptides **34**-**36** and can be rationalised as a combination of them. The spectrum shows two minima; the first one close to the 220 nm region but shifted towards 210 nm relative to the match decamers, and a second one in the 200 nm region but shifted to red and weaker than that displayed by mismatch **36**. This was interpreted as indicative of mismatch **38** adopting both random coil and folded hairpin structures.[99]

Figure 3.21 shows the absorbance spectra for the four peptides, the plots overlap throughout the spectra. The concentration of each peptide was calculated using the 267 nm wavelength (see section 7.2.8) and was in the 101-114 μM range.

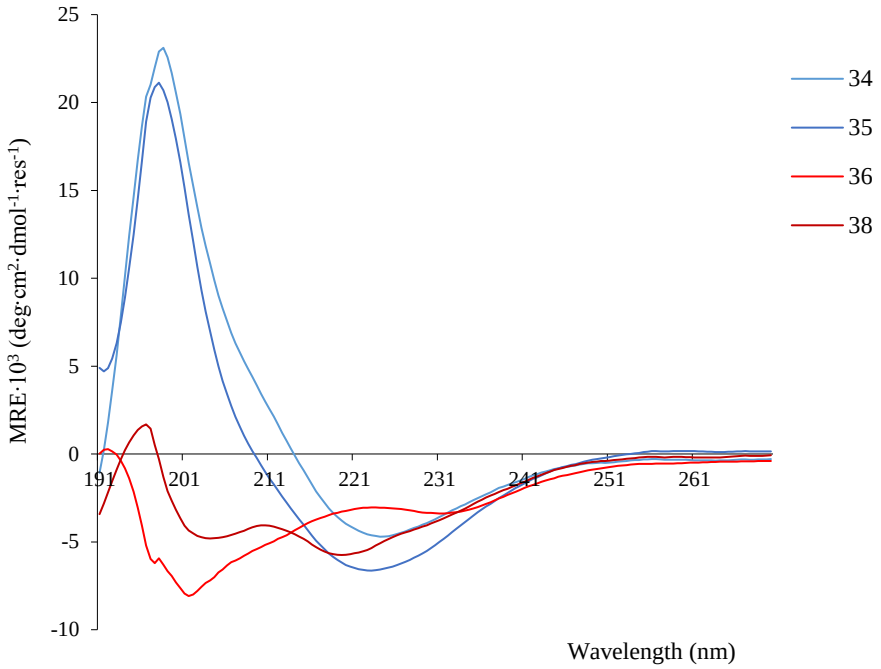


Figure 3.20: Mean residue ellipticity of peptides **34**-**38** at 25 °C. Spectra are reported for ~100 µM peptide samples in methanol (see section 8.2).

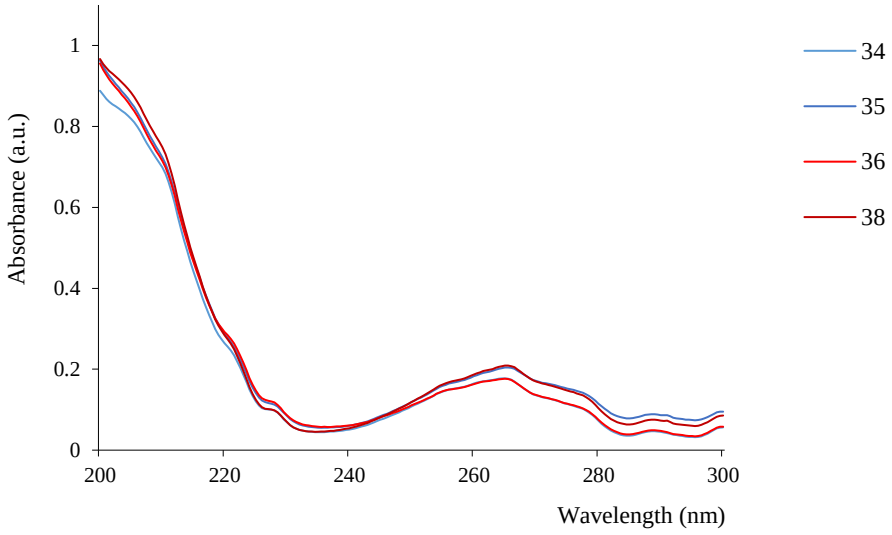
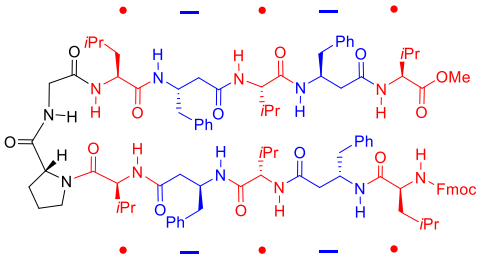


Figure 3.21: Absorbance spectra of peptides **34**-**38** showing consistent overlap, concentrations range from 101 to 114 µM in methanol (see section 8.2).

These results were in agreement with the previous NMR analyses, suggesting that peptides bearing strands of matching α/β-residues formed stable folded structures while peptides with mismatching α/β-residues could not do this to the same extent.

3.4. A dodecamer α/β-peptide β-hairpin

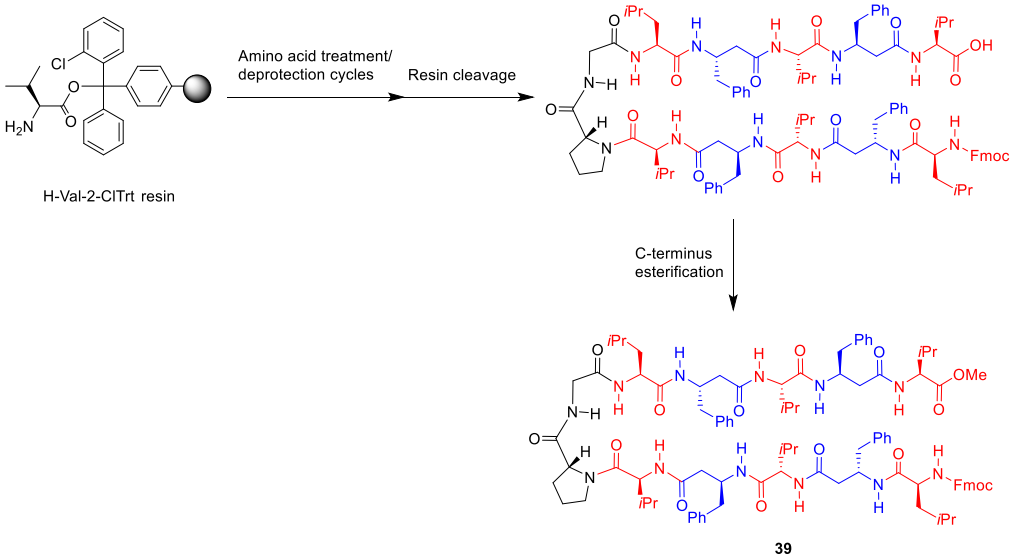
Having completed the study of our original set of decamer α/β-peptides we were interested in probing the boundaries of our PMC system regarding how much information could be encoded in the peptides (i.e. how long can their α/β-residue sequence be) while keeping control of their conformation. Thus, we set out to synthesise a longer β-hairpin containing an extra pair of β-residues (Figure 3.22). Matched dodecamer **39** is an analogue of matches **34** and **35** bearing four β^3hPhe residues alternated with α-residues in the sequence.



39

Figure 3.22. Dodecapeptide **39** bearing matching α/β-strands.

The dodecamer was synthesised using a standard manual Fmoc-based SPPS methodology to yield the C-terminal free carboxylic acid. This was then converted into the methyl ester derivative using the same methodology used with the decamers (Scheme 3.4). The Fmoc-deprotection step on the N-terminal residue was omitted, and the peptide was purified through HPLC (methanol/water). The identity and purity of the peptide was assessed by MS and analytical HPLC.



39

Scheme 3.4: Overview of the synthesis of dodecamer **39** (for full conditions see sections 7.2.1 and 7.3.2).

The solubility of dodecamer **39** was notably reduced compared to the decamer analogues, which difficulted its purification and characterization. A ^{1}H NMR spectrum was obtained in CD_3OH, Figure 3.23 shows a very dispersed group of signals in the amide region and a distinctive group of aliphatic and aromatic groups. However, the solution saturated at very low concentration (~100 µM) and key 2D NMR experiments were unfruitful at this low concentration, thus full assignment of the spectrum was precluded. Unfortunately, unlike matched decamers **34** and **35**, dodecamer **39** was not soluble in chloroform so conformational studies based on NMR could not be carried out.

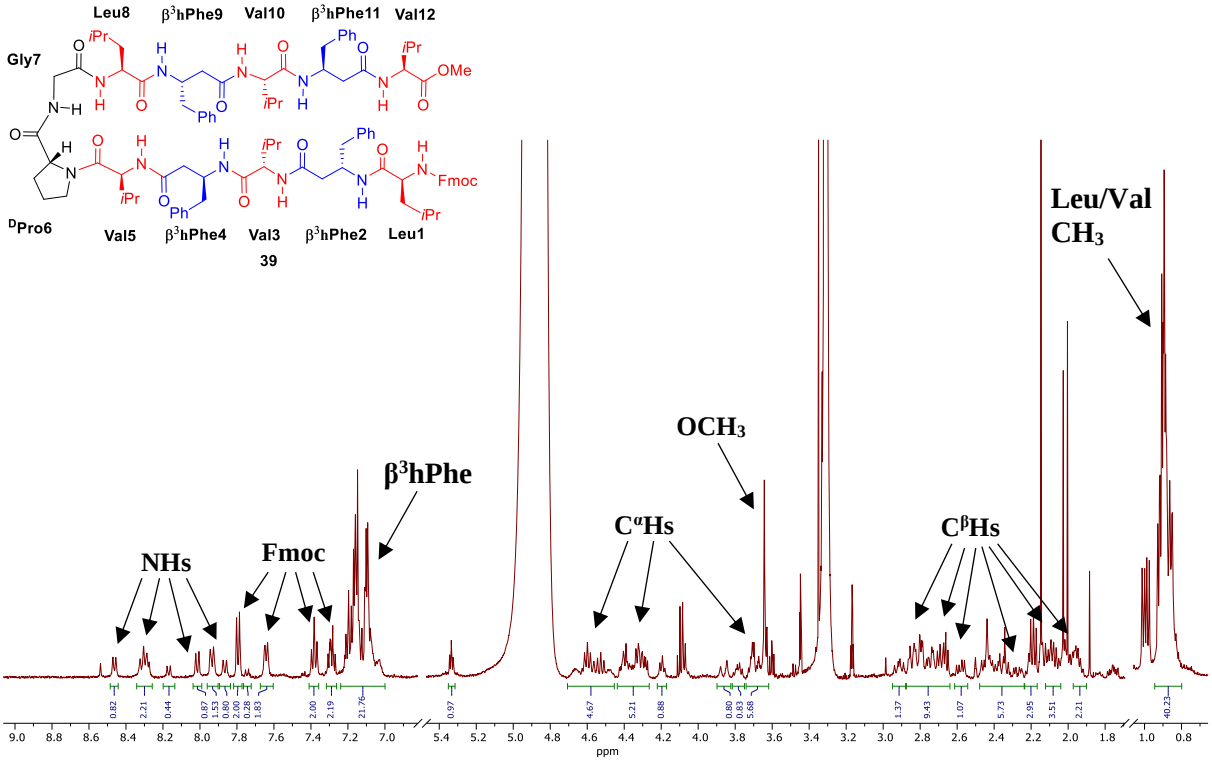


Figure 3.23: 500 MHz ^{1}H NMR spectrum of peptide **39** (~100 µM) in CD_3OH.

On the other hand, given that the peptide did present partial solubility in methanol, we proceeded to analyse it by CD, since this technique allows analysis of considerably diluted species. in Figure 3.24 shows the CD spectrum of dodecamer **39** at a 47 µM concentration in methanol, for comparison, the spectrum of decamer **35** at an analogous concentration (49 µM) is shown too. The dodecamer presented the same pattern as **35**; a minimum at 220 nm and a maximum at 200 nm, characteristic of a β-hairpin conformation. As was the case for matched peptide **34**, the weaker intensity of the band close to 220 nm points to a smaller population of β-sheet in the dodecamer. This can be rationalised as the result of the introduction of an additional pair of flexible methylene (form the β^{3}hPhe residues), which could destabilise the sheet-like antiparallel strands.[96]

Because of the design of dodecamer **39**, the synthesis of an isomeric analogue with mismatching α/β-strands (with alternating α/β-residues) would present difficulties (the effect of having β-residues in the

N-terminus or *i* position of the turn had not yet been studied in our designs). Therefore, it was decided to focus on longer peptide sequences to evaluate the effect of match *vs* mismatch α/β-strands in secondary structure.

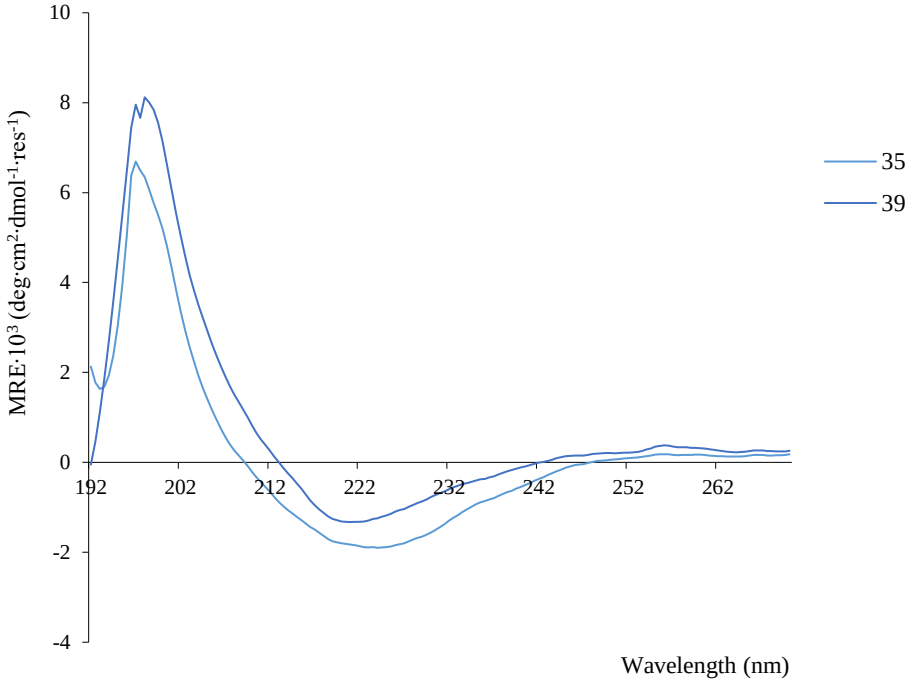


Figure 3.24: Mean residue ellipticity of peptides **39** and **35** at 25 °C. Spectra are reported for ~50 μM peptide samples in methanol.

3.5. A series of tetradecamer α/β-peptide β-hairpins

3.5.1. Acetyl-capped tetradecamers

We envisioned that tetradecamers with 6-residue strands would allow us to obtain sets of match/mismatch isomers containing four β-residues without the potential issues posed by using shorter strands. Moreover, the diminished solubility displayed by dodecamer **39** prompted us to rid the design from the aromatic, bulky N-terminus Fmoc group. We were concerned it could be inducing intermolecular interactions leading to aggregation. Thus, the pair of tetradecamers **40** and **41** was synthesised acetylating the N-terminus instead. (Figure 3.25). An additional pair of α-residues were inserted to yield the new set of peptides; for matched peptide **40** the residues were positioned close to the β-turn and for mismatch **41** the residues were positioned so that the β-residues are not directly facing each other in the strands.

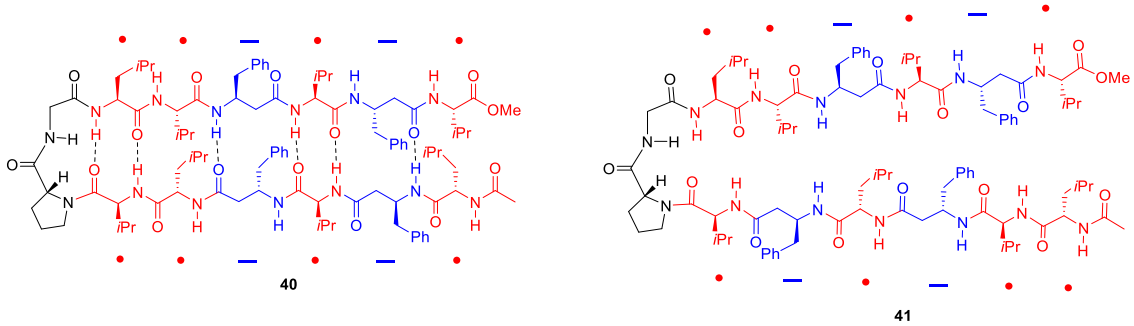
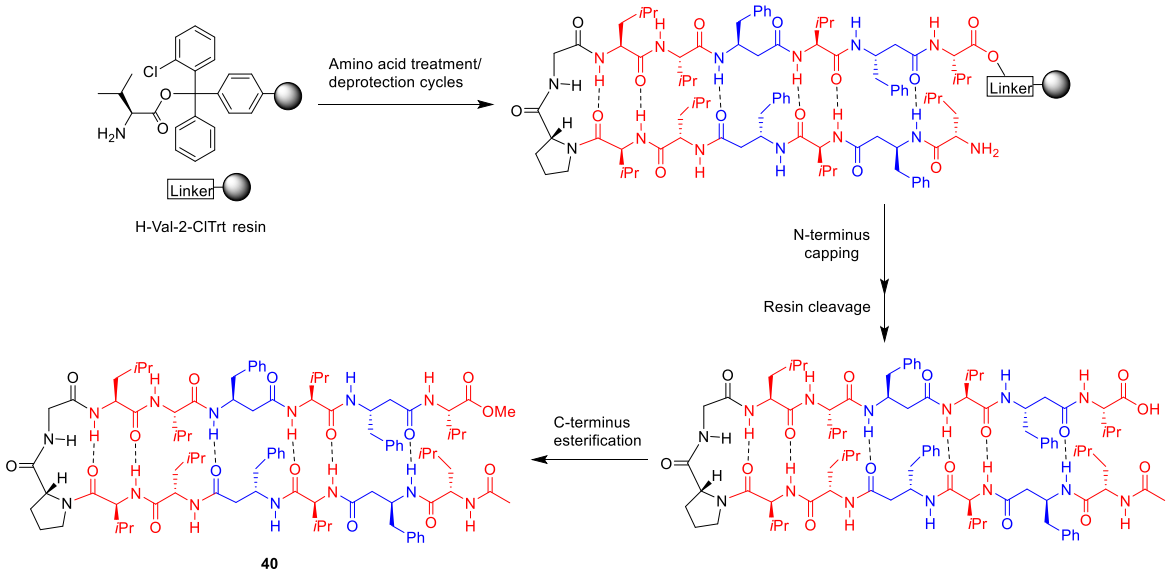


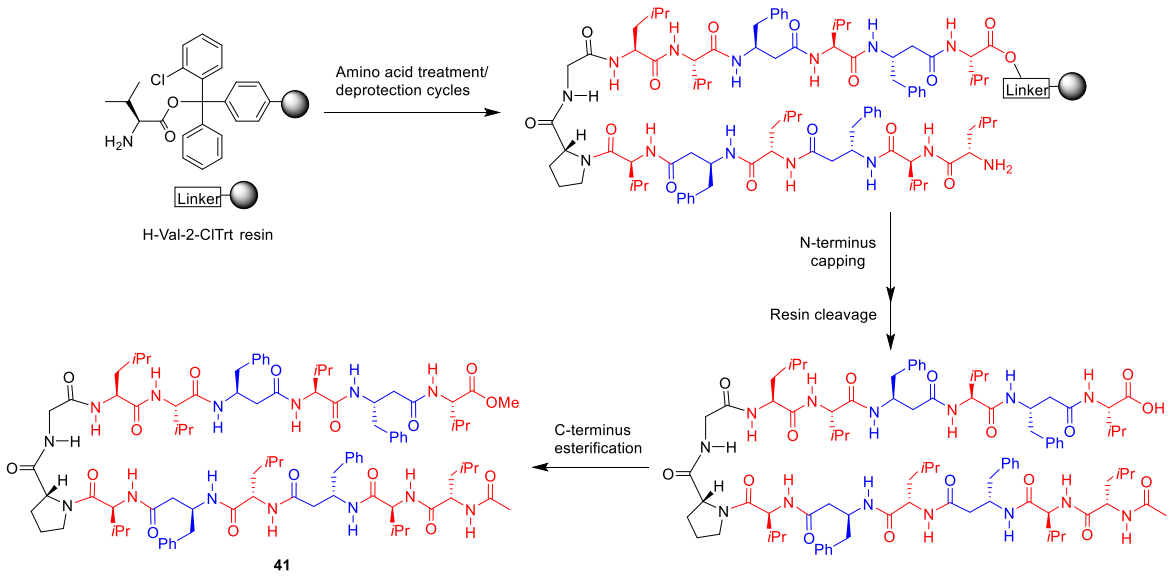
Figure 3.25: Tetradecapeptides **40** and **41** bearing matching and mismatching α/β-strands respectively.

The tetradecamers were prepared using a standard manual Fmoc-based SPPS methodology to yield the C-terminal free carboxylic acids. These were then converted into the appropriate methyl ester derivative (Schemes 3.5 and 3.6). Prior to resin cleavage, the terminal Fmoc protecting group was removed and the N-terminus was acetylated. The peptides were purified through HPLC (methanol/water), and the identity and purity of the peptides was assessed by MS and analytical HPLC.

Unfortunately, the solubility of the tetradecamers was very poor compared with the peptide hairpins we had worked with so far. Even HPLC purification and analysis became very challenging, requiring multiple runs to obtain pure peptide in the nanogram scale. NMR analysis of these compounds was precluded, but CD studies could still be carried out.



Scheme 3.5: Synthetic route of tetradecamer **40** (for full conditions see sections 7.2.1 and 7.3.2).



Scheme 3.6: Overview of the synthesis of tetradecamer **41** (for full conditions see sections 7.2.1 and 7.3.2).

Figure 3.26 shows the CD spectra for the tetradecamers at ~10 µM concentrations in methanol. Match **40** displayed an anomalous CD spectrum with an intense maximum at 206 nm and a minimum at 229 nm. Both bands exhibit a shift to the red with respect to the bands seen for the previous matches. However, peptides with type II and II' β-turns presenting considerable β-sheet content have been shown to display bands at these wavelengths.[94,96,100] It has also been reported that modest variations in the dihedral angles in the region of the turn cause changes in the CD spectrum, but that the general pattern remains similar to that of the β-sheet CD spectrum with the bands shifted to the red.[101]

On the other hand, the spectrum of mismatch analogue **41** presented a pattern closer to what had been observed for mismatch **38**; a minimum close to the 220 nm and a modest maximum close to 198 nm suggesting the onset of a β-sheet structure, but also a negative shoulder near the 205 nm region, consistent with some random coil content.

It was clear that further studies that could shed more light on the conformation adopted by these molecules in solution would be very difficult due to their poor solubility. Moreover, further studies on our ability to control conformation through the use of match *vs* mismatch sequences (the PMC concept) in even longer, more complex designs would be limited by this same issue.

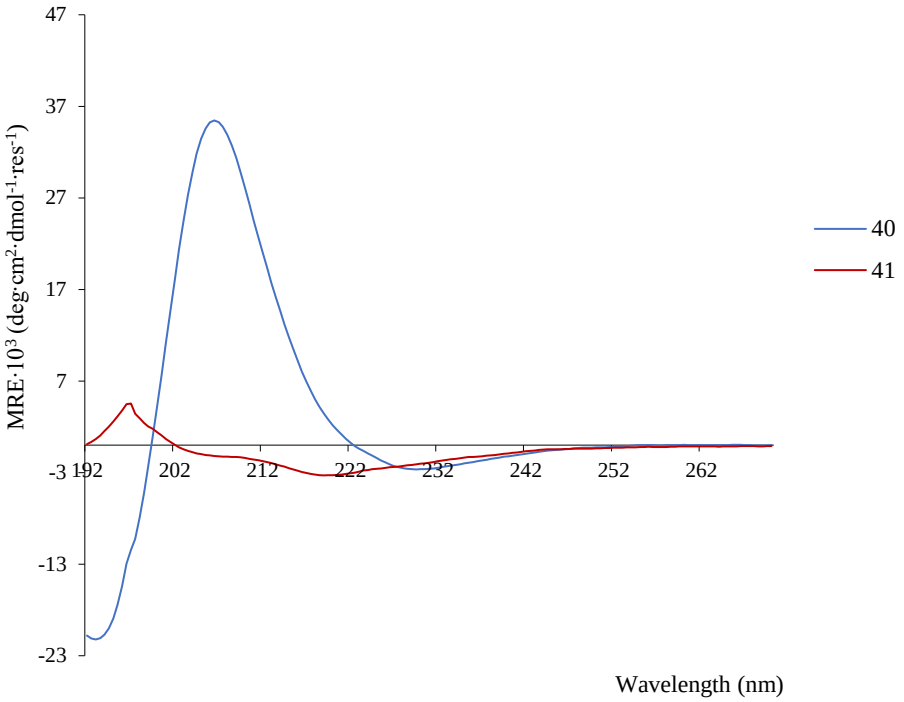


Figure 3.26: Mean residue ellipticity of peptides **40** and **41** at 25 °C. Spectra are reported for ~10 μM peptide samples in methanol.

3.5.2. Tetradecamer β-hairpins bearing solubilizing groups

To circumvent the low solubility problem, we came up with two modifications to our original design. On the one hand, we envisioned that incorporating polar residues into the peptides could improve their solubility in polar organic solvents. Threonine (bearing a polar OH group) was chosen to this effect, given its propensity towards β-structures because of branching in the side-chain.[102] Thus, tetradecamer match **42** was designed (Figure 3.27, top), an analogue of **40**, where four Thr residues have been introduced, replacing Val and Leu residues in the original peptide, and are positioned so that any obvious direct cross-strand side-chain interactions that could be determinant for the peptide adopting a folded structure are avoided. The methyl ester in the C-terminus was replaced for an amide, with the purpose of avoiding the post-SPPS modification, which presented difficulties for the previous tetradecamers.

Secondly, given that the use of poly(ethylene glycol) PEG-peptide conjugates has been shown to enhance solubility of hydrophobic peptides in both aqueous and polar organic solvents.[103,104] Peptide **43** was designed, where $(EG)_3$ chains have been incorporated in both the N- and the C-terminus (Figure 3.27, bottom).

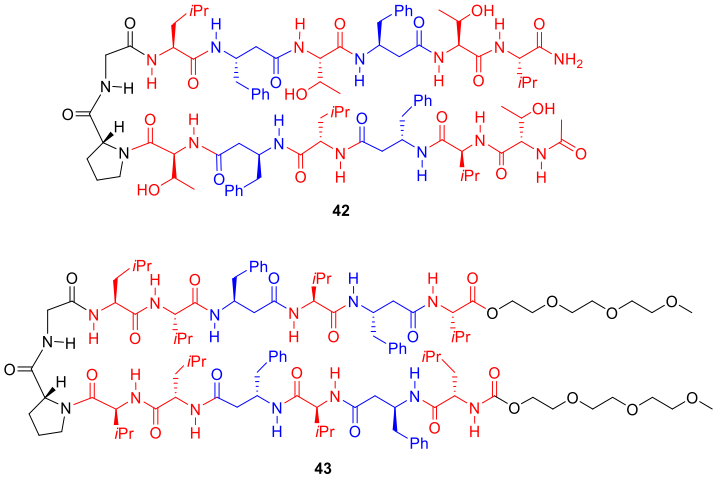
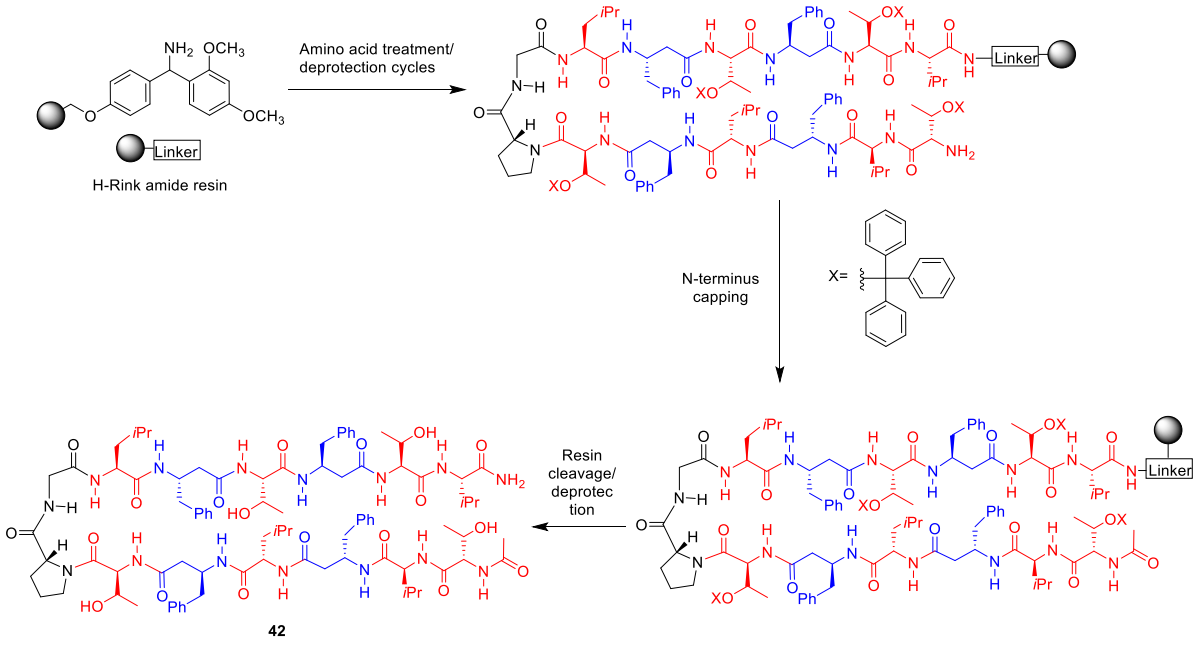
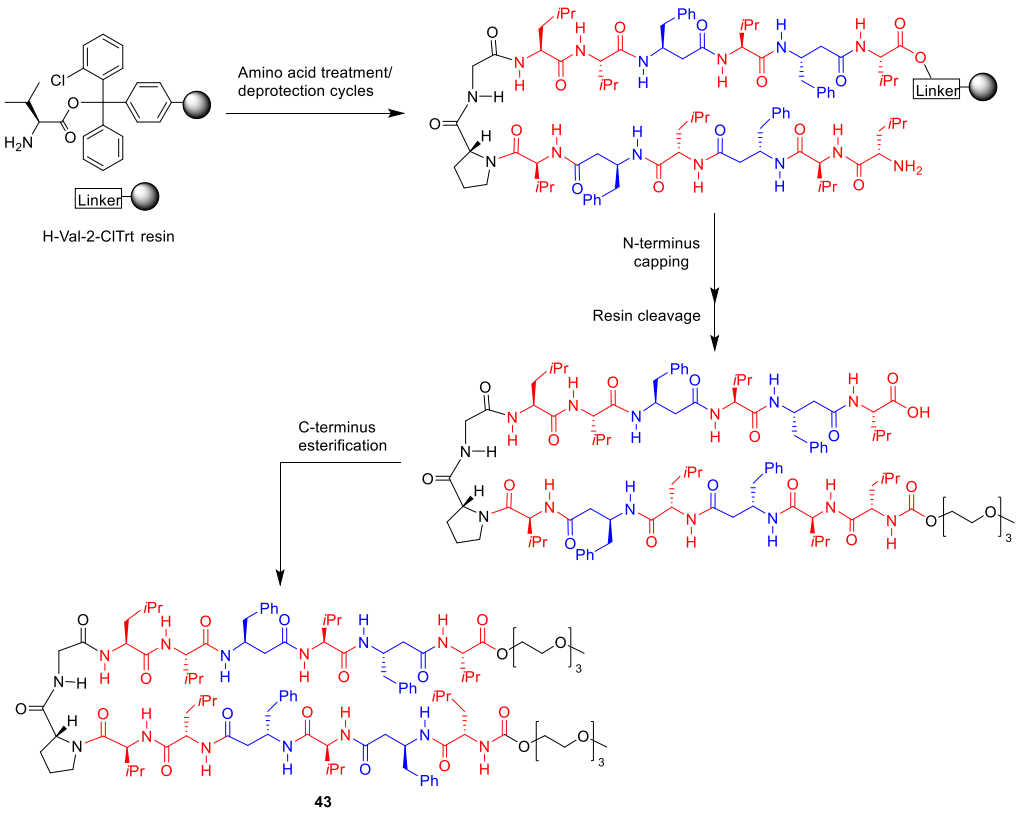


Figure 3.27: α/β-tetradecapeptides **42** (**top**) and **43** (**bottom**) bearing Thr residues and $(EG)_3$ capping groups, respectively.

Peptide **42** was synthesised using a standard manual Fmoc-based SPPS methodology on a H-Rink-amide resin to yield the C-terminal amide. Prior to resin cleavage, the terminal Fmoc protecting group was removed and the N-terminus was acetylated (Scheme 3.7). In the case of peptide **43**, a 2-ClTrt resin was used to yield the C-terminal free carboxylic acid. Prior to resin cleavage, the terminal Fmoc protecting group was removed and the peptide was treated with $(EG)_3$-succinimidyl carbonate and DIPEA to yield the N-terminal $(EG)_3$ carbamate. The peptide was then cleaved from the resin and the C-terminal carboxylic acid was converted to the $(EG)_3$ ester by treating it with triethylene glycol monomethyl ether containing HCl (5% v/v) at 60 °C for 48 hours (Scheme 3.8, see section 7.3.4).The peptides were purified through HPLC and (methanol/water) the identity and purity of the peptides was assessed by MS and analytical HPLC.



Scheme 3.7: Overview of the synthesis of tetradecamer **42** (for full conditions see section 7.2.1).



Scheme 3.8: Overview of the synthesis of tetradecamer **43** (for full conditions see sections 7.2.1 and 7.3.4).

Unfortunately, the solubility of both of the new modified tetradecamers did not improve. HPLC purification and analysis was still challenging and in the case of tetradecamer **43** isolation of the peptide from the esterification reaction mixture was not easy. A variety of solvent mixtures (organic and aqueous) were tried with both peptides but we never managed to get them in solution to perform NMR analysis.

However, diluted solutions in methanol could still be obtained, and so CD spectra of both peptides were recorded. Figure 3.28 shows the spectra for the modified peptides along with the original tetradecamer (**40**). It can be seen that all three peptides display a very similar pattern. Threonine analogue **42** shows a maximum at 203 nm and a minimum at 222 nm, both bands shifted back towards blue with respect to **40** and closer to those displayed by the matched decamers and dodecamer **39**. In the case of PEG derivative **43**, an intense maximum was seen at 205 nm and a minimum at 227 nm, similar to what was observed for tetradecamer **40**, the higher intensity of both bands suggests a larger population of β-sheet secondary structure, potentially promoted by the $(EG)_3$ chains.

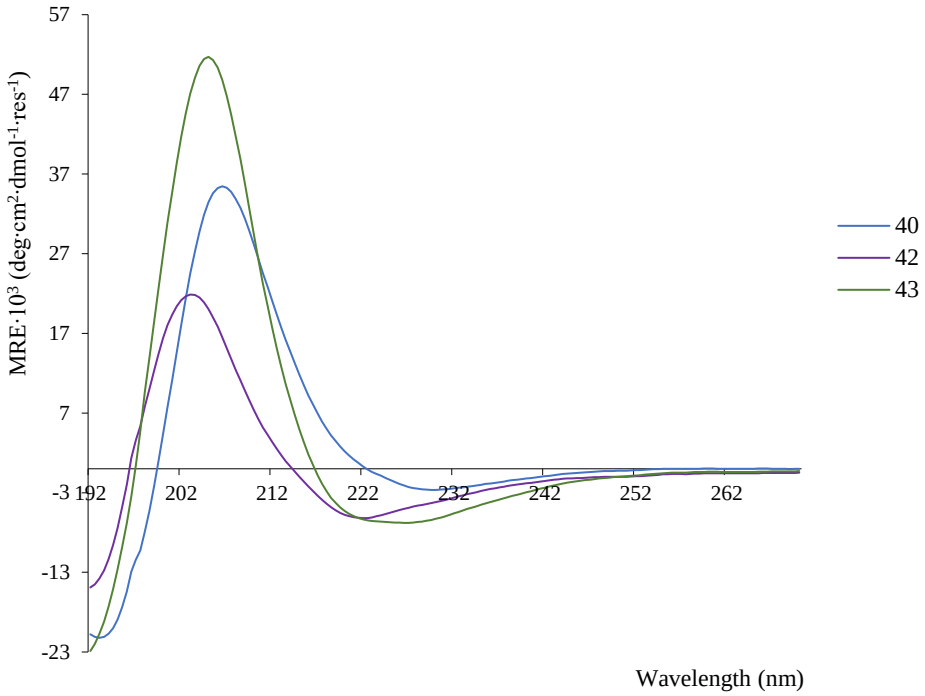


Figure 3.28: Mean residue ellipticity of peptides **40**, **42** and **43** at 25 °C. Spectra are reported for ~10 μM peptide samples in methanol.

3.6 Conclusions and future work

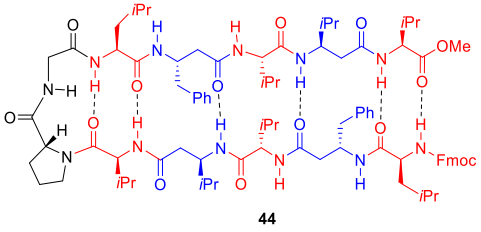
A set of five α/β-peptide decamers were synthesised to provide an initial demonstration of the divergent behaviour of α/β-peptide hybrids bearing matching *vs* mismatching α/β-strands. NMR studies of decamers **34** and **35** in chloroform (and methanol in the case of **34**), bearing matching α/β-strands, gave strong proof of the peptides adopting a β-hairpin structure supported by intramolecular H-bonds between the facing strands, cross-strand interactions were evidenced by NOEs. Concentration dependence of chemical shift studies of **35** in chloroform also suggested the formation of infinite β-sheet-like aggregates at high concentrations. CD studies in methanol confirmed these observations, both peptides exhibited signature spectra for the β-hairpin structure.

NMR studies of mismatch **36** in methanol suggested formation of the β-turn at the ^{D}Pro-Gly segment, but cross-strand NOEs were not observed, which was incompatible with a β-hairpin conformation. CD analysis in methanol confirmed that peptide **36** existed predominantly as a random coil structure. Unfortunately, NMR studies of **36** in chloroform could not be carried out due to poor solubility. As a general observation, decamer mismatches presented reduced solubility relative to matches, which could be associated with them adopting disordered conformations. NMR analysis of mismatch **38** in chloroform also suggested the formation of a β-turn at the ^{D}Pro-Gly segment. However, critical cross-strand NOEs were still absent. This and other NMR observations suggested the peptide was structured

in the β-turn region but less so as the strands progressed to the termini due to fraying. CD was in agreement with this hypothesis since the spectra of **38** showed the peptide adopted both random coil and folded β-hairpin structures.

As the synthesis of longer α/β-peptides containing more β^{3}hPhe residues was attempted, their solubility became restrictive. NMR analysis was precluded. Circular dichroism analysis of a dodecamer bearing four β^{3}hPhe at matching positions in the strands (**39**) presented a signature spectrum for a β-hairpin similar to that obtained for the decamer matches. A pair of matched and mismatched tetradecamers **40** and **41** was synthesised and analysed by CD too. Match **40** yielded an anomalous spectrum, with the same pattern of bands as the other matches but shifted toward red, which was interpreted as a change in the dihedral angles of the turn, where a significant population of β-sheet structure is present. On the other hand, mismatch **41** presented a CD spectrum suggesting the presence of both disordered and β-hairpin structures.

Modifications to the α/β-peptides were attempted using polar amino acid residues and solubilizing capping-groups, but the solubility of the long designs could not be improved. Future work in this project should focus in the design soluble longer α/β-peptides. To this respect, a successful approach currently taken in the group is substituting the β^{3}hPhe residues for β^{3}hVal. Dodecamer **44** bearing four β-residues including two β^{3}hVal (designed and synthesised by Dr Jinqiao Dong) has shown an improved behaviour in solution which has allowed NMR analysis pointing to a folded β-hairpin conformation (Figure 3.29).



44

Figure 3.29: Structure of dodecamer **44** bearing matching α/β-strands, with a pair of β^{3}hPhe and a pair of β^{3}hVal.

In summary, our studies on this first series of α/β-peptide hybrids point to a sequence-selective assembly of folded β-hairpins, where peptides bearing α/β-residues arranged in matching positions in the facing strands will form stable intramolecular antiparallel β-sheets, whereas strands bearing mismatching sequences of α/β-residues will not, showing the presence of disordered structures. This provides evidence for peptide Morse code to serve as an information-coding system.

Chapter 4. Design of an organic media-soluble PMC replicator

The work described in this chapter was designed by the author of this thesis (AAC) and Prof. Anthony Davis. Some of the work discussed in this chapter was carried out in collaboration with Danny Burke and was reported in his thesis towards an MSci degree at the University of Bristol. Specific contributions are mentioned in the text and in sections 7 and 8. All other experiments and analyses were carried out by AAC.

4.1. Project aims

Chapter 2 describes the problems we encountered while studying self-replication in peptides that form amphiphilic β-sheet aggregates. These consisted mainly of our inability to analyse the kinetics of peptide fragment ligation due to unresolvable mixtures and the instability of the reacting species in the aqueous medium employed. Even Ashkenasy's literature system, employing only α-amino acids, gave irreproducible results. These issues kept us from probing ISR in α/β-peptide hybrids. On the other hand, given our success in demonstrating the sequence-selective assembly of short β-hairpins in organic media (described in chapter 3), we envisioned to adapt a self-replicating system from these hydrophobic α/β-peptide variants to study ISR, a key feature of PMC. By moving to organic media, we could avoid some of the problems associated with the aqueous medium, specifically the requirement for Cys (liable to oxidation) in the nucleophilic component and the vulnerability of the electrophile to hydrolysis. We also decided to use selectivity, rather than kinetic measurements, as a test for ISR.

Therefore, we devised a simple system based on our previous work; it consisted of two short linear template peptides, one bearing a palindromic sequence (**45**) and another a non-palindromic sequence of α/β-residues (**46**, Figure 4.1). Firstly, the structure of this set of molecules would be assessed in order to extend our results from chapter 3 to the intermolecular sequence-selective formation of β-sheet-like structures. Figure 4.1 shows the antiparallel β-sheet arrangement expected for palindrome **45**, where all the H-bonding sites engaged throughout the strands. On the other hand, non-palindromic **46**, which is not self-complementary, was not expected to assemble into β-sheets as efficiently by itself. The figure also shows the misaligned arrangement of peptide **46** with a pair of H-bond acceptors uncoupled in the middle of the β-strand.

Once α/β-sequence based recognition in this system was established, template-directed formation of covalent bonds would be evaluated with these hydrophobic variants using similar experiments to those attempted with the amphiphilic peptide variant in chapter 2: fragments of the templates would be synthesised and the ligation of those fragments analysed. We expected self-complementary peptide **45** would recognise and template the coupling of its corresponding fragments while non-palindromic **46**

should not, so coupling of its corresponding fragments should proceed linearly (Figure 4.2). This would support PMC information transfer.

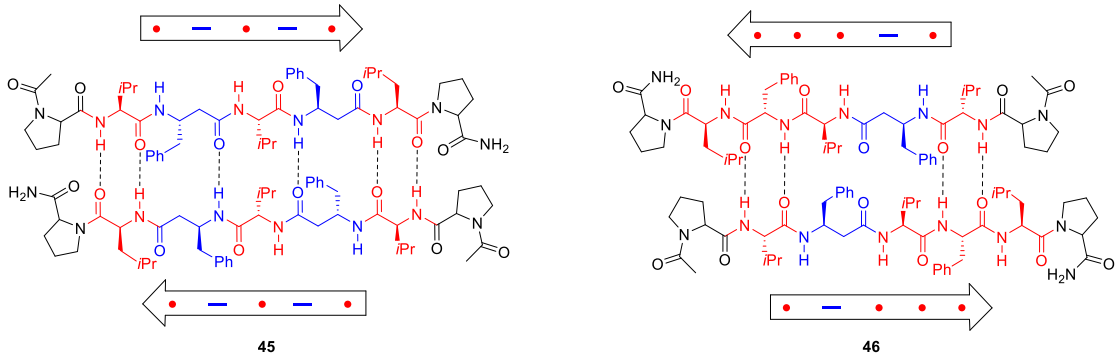


Figure 4.1: Structure of α/β-peptide hybrid 7-mers **45** and **46**. 7-mers assembled as β-sheets, self-complementary peptide **45** is shown to adopt an ordered antiparallel arrangement (**left**), whereas non-palindromic analogue **46** is shown adopting a less ordered, less favoured arrangement (**right**).

4.2. Peptide design

The pair of short linear heptamer α/β-peptides **45** and **46** (Figure 4.1) were built from the same hydrophobic residues used for our structural studies: Leu, Val and β^{3}hPhe. Palindrome **45** carried a pair of β^{3}hPhe residues while in non-palindromic **46** one of the β-residues was replaced by the α-analogue Phe. The terminal β-sheet breaking Pro residues featured in Ashkenasy[48] and Gellman's[105] amphiphilic designs were also incorporated. Constituting fragments of the templates corresponding to the breakage of the amide bond at the Val4 residue were also built. These yielded three fragments: two free-NH_2 nucleophilic fragments **47** and **48**, and a common electrophilic fragment **49** functionalised as an *N*-hydroxysuccinimide ester (Figure 4.2). NCL was avoided in this system, given the complications we had encountered previously using it as coupling method (see chapter 2) and since other options were available as aqueous media was avoided. Hence, we opted for an activated *N*-hydroxysuccinimide ester given its widespread use in peptide coupling and conjugation.[104,106] Figure 4.2 shows the template-assisted coupling of fragments **47** and **49** by peptide **45**, contrasting with the less efficient recognition between non-palindromic **46** and its constituting fragments.

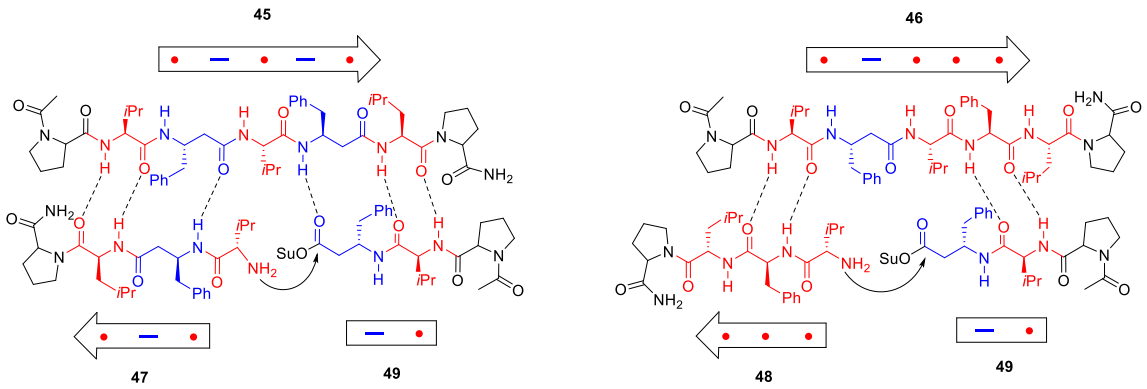
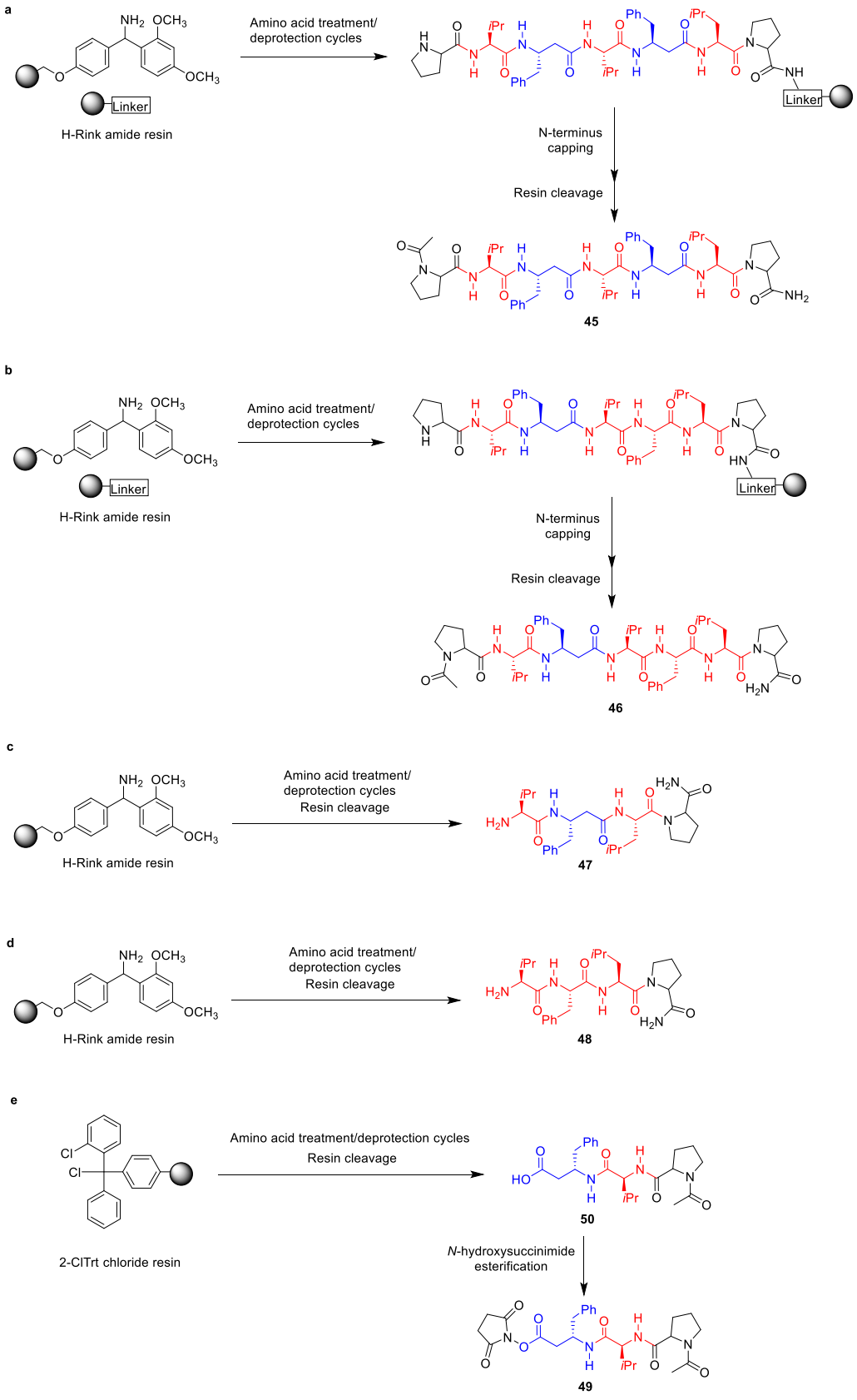


Figure 4.2: **Left**. Coupling of fragments **47** and **49** templated by peptide **45**. **Right**. Inefficient recognition between peptide **46** and corresponding fragments (**48** and **49**). Su = succinimide.

4.3. Peptide synthesis

The two templates and three fragments were synthesised using a standard Fmoc-based manual solid-phase peptide synthesis (SPPS) methodology (Scheme 4.1 see also sections 7.2.1 and 7.3.3). An H-Rink amide resin was used to obtain peptides **45**, **46**, **47** and **48**. In the case of the two templates the N-terminus was acetylated and for the nucleophilic fragments, the terminal Fmoc protecting group was removed. The electrophilic fragment was synthesised as a C-terminal carboxylic acid (**50**) using a 2-chlorotrityl chloride resin, the N-terminus was acetylated prior to resin cleavage. After cleavage, the free acid at the C-terminus (**50** was used without further purification) was converted to an *N*-hydroxysuccinimide ester (**49**), following a literature procedure of a similar transformation (see section 7.3.3 for full details).[107] The peptides were purified by preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)). The identity and purity of the peptide fragments was assessed by ^{1}H and ^{13}C NMR (see section 7.6.1), ESI MS and analytical HPLC. In the case of the templates, MALDI-TOF MS and analytical HPLC was used (see section 8.1.3; synthesis and characterization of peptides **45-50** was first carried out in collaboration with Danny Burke).



Scheme 4.1: Overview of the synthesis of peptides **45** (**a**), **46** (**b**), **47** (**c**), **48** (**d**) and **49** (**e**; for full conditions see sections 7.2.1 and 7.3.3).

4.4. Structural study of heptamers 45 and 46

A brief study of the secondary structure adopted by the two 7-mer peptide hybrids was carried out by analysing their CD spectra (Figure 4.3, spectra recorded by Dany Burke). We emulated the conditions used for our study of the β-hairpin decamers, so the spectra were recorded at ~100 μM in methanol (see section 3.3.6). The spectra displayed by palindrome **45** bears similarity to that of matched hairpins **34** and **35**, a minimum at 220 nm and a modest maximum at 198 nm were observed, these bands correlated with the peptide adopting a β-sheet secondary structure. On the other hand, non-palindromic peptide **46** displayed an intense negative band at 204 nm with no clear minimum at 220 nm, similar to the spectra of mismatch **36**, consistent with the peptide adopting predominantly a random conformation.

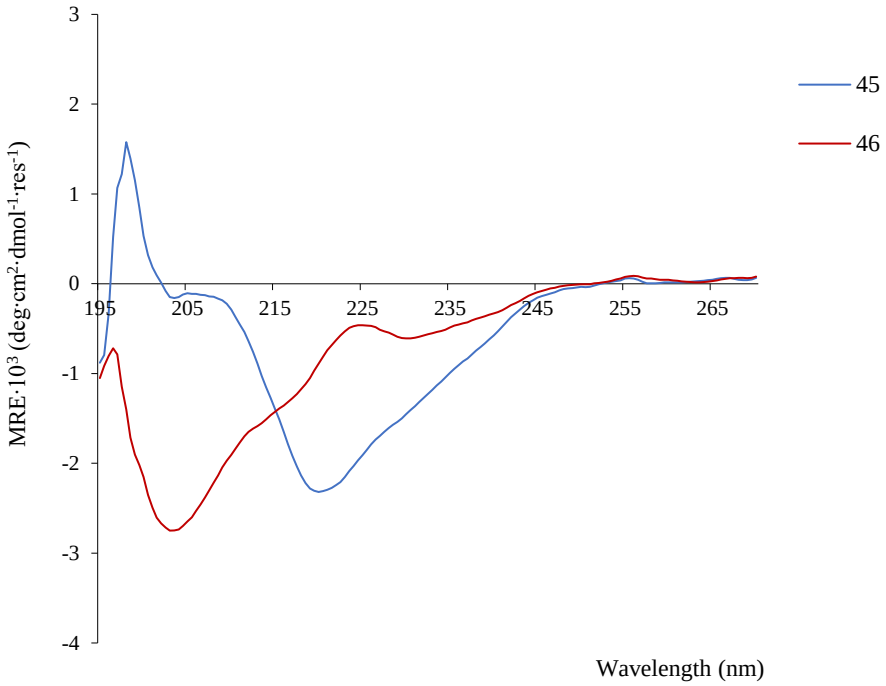
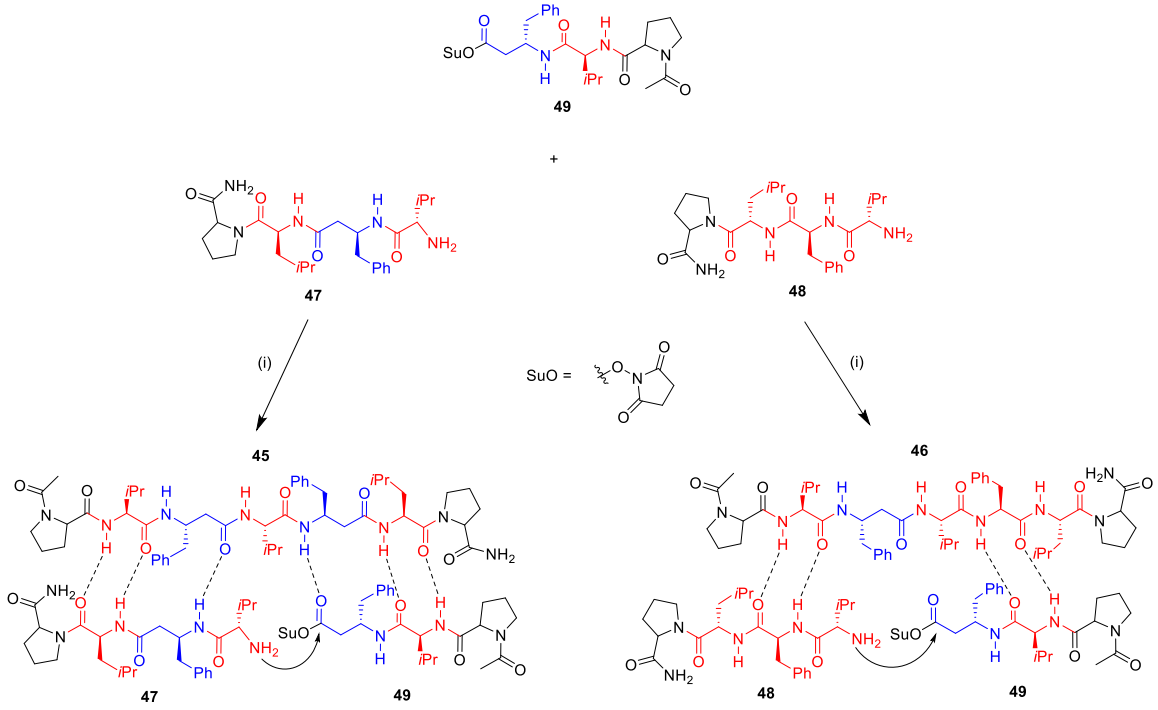


Figure 4.3: Mean residue ellipticity of peptides **45** and **46** at 25 °C. Spectra are reported for ~100 μM peptide samples in methanol.

4.5. Competing coupling experiments

Given our experience with monitoring ligation kinetics in the amphiphilic systems (see chapter 2) and having realised the difficulties these studies entail, we envisioned a different approach, which consisted of a competition experiment. Since we expected that the coupling of the fragments corresponding to the self-complementary template **45** would proceed autocatalytically while the coupling of the fragments corresponding to non-palindromic **46** would not, this behaviour should be reflected in an enrichment of **45** over **46** when both coupling reactions compete with each other. Therefore, if equimolar amounts of

fragments **47** and **48** were mixed with a limiting amount of electrophile **50**, we would expect **45** to be produced faster and in larger yields compared to **46** (Scheme 4.2). This would provide a first indication of informational self-replication in our system.



Scheme 4.2: **Top**. Coupling of a mixture of the nucleophilic fragments **47** and **48** with activated ester **49** to generate heptamers **45** and **46**. **Bottom**. The generated product **45** can template further coupling of its constituting fragments. In contrast, the generated non-palindromic **46** is shown to recognise its constituting fragments poorly. (i) Equimolar amounts of the three fragments (1 mM) in methanol, 2 equivalents of DIPEA.

Reaction conditions for the successful coupling of the nucleophilic fragments **47** and **48** to the electrophilic fragment **49** were developed (see section 7.5.1 for full details). Given that our system was designed to adopt a β-sheet secondary structure in organic media we attempted the coupling of the peptide fragments in methanol. Initially, one pot coupling of the fragments in methanol in the presence of 2 equivalents of base (DIPEA) produced the template products in low yield due to the fast hydrolysis of the *N*-hydroxysuccinimide ester in the basic methanolic media. This was circumvented by premixing the nucleophilic fragments with the base and then adding the activated ester to this mixture. An analytical method to quantify the products from the coupling reactions was also developed (see section 7.5.1 for full analytical conditions). The method consisted of a relatively long analytical HPLC run combining step gradients and isocratic elution.

Figure 4.4 shows representative HPLC traces of the coupling competitions, broad peaks with t_R = 35.2 and 38.3 min correspond to heptamers **45** and **46**, respectively. The ratio of the heptamers in the mixture was determined by comparing the area under the peaks corresponding to both peptides. It was found that the ratio of **45** to **46** under the standard reaction conditions was 62:38, close to a 2:1 relationship.

This ratio was found to remain unvaried over the course of the reaction, for samples monitored after 1 and 4 hours of initiating the reaction.

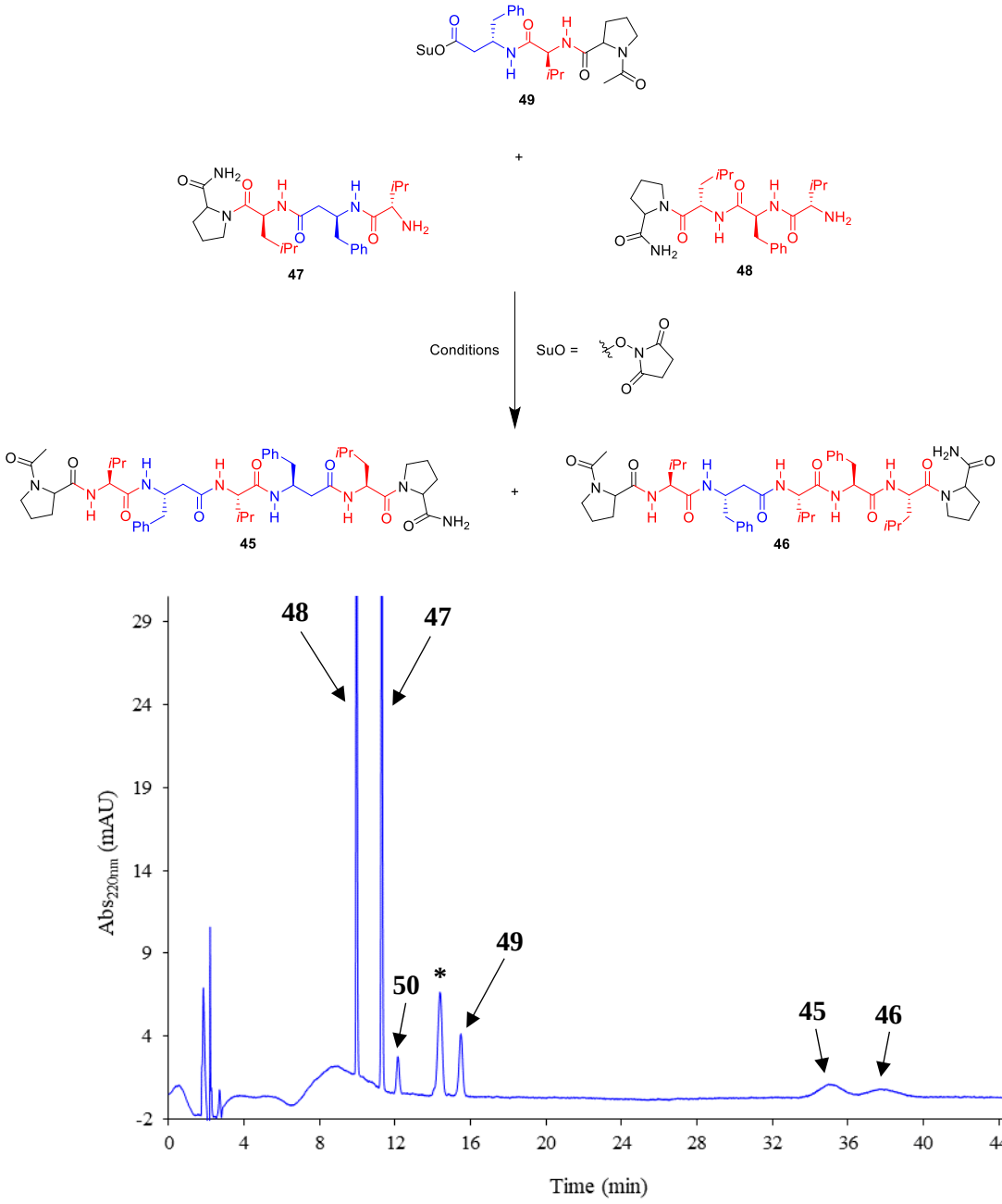
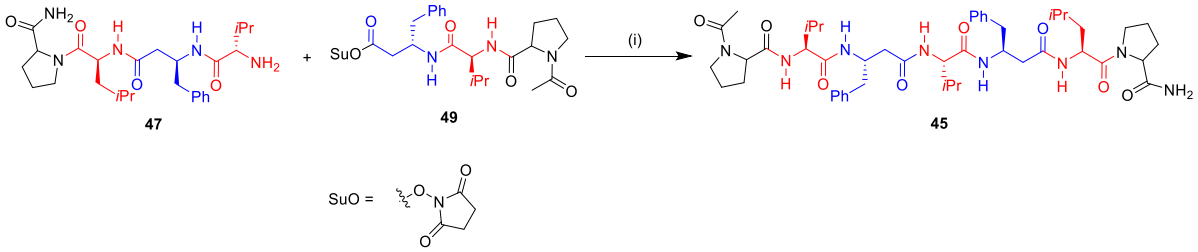


Figure 4.4: **Top**. Scheme of the coupling of fragments **47** and **48** (equimolar) with fragment **49**. **Bottom.** Representative HPLC traces of the ligation of fragments **47** and **48** with **49** for a reaction that proceeded for 1 h (for full reaction and analytical conditions see section 7.5.1). * denotes a methyl ester derivative of **50** (free-carboxylic acid precursor) as detected by LC-MS.

Encouraged by this selectivity, we decided to gain further proof of an autocatalytic behaviour in the synthesis of **45**. Therefore, we set out to evaluate the coupling of **47** to **49** in reactions seeded with template **45**.

4.6. Kinetic studies of the template-assisted synthesis of heptamer 45

We studied the rate of the coupling of fragment **47** to electrophile **49** when template **45** was externally added to the reaction (Scheme 4.3). We expected that the seeding of **45** would affect the initial rate of the fragment coupling, which would point to a template-directed self-replication.



Scheme 4.3: **Top**. Coupling of fragments **47** and **49** to generate heptamer **45**. (i) **45** at different concentrations (~100 and ~250 µM), 2 equivalents of DIPEA.

For these studies, an analytical method to quantify heptamer **45** was optimised. Considering the difficulties we had experienced with the recovery of peptide **1** (chapter 2) and given that these analyses were being carried out in an HPLC equipped with a manual injection port (as opposed to an automated injector as used for the analysis of **1**) which could introduce an additional error source from volume variation, we quantified peptide **45** by using a calibration plot based on an internal standard **Std** (see section 7.5.2). Thus, throughout these studies, 80 µL samples of the ligation reactions were taken at different time points and mixed with an equal volume of a 250µM solution of peptide **Std**. By doing this, the concentration of the internal standard was kept constant in every sample and the ratio of the areas under the curve of the peaks corresponding to **45** and **Std** could be used to calculate the concentration of heptamer **45** at each sample point (for full details see section 7.5.2).

For our first attempts at evaluating the autocatalytic effect in the synthesis of heptamer **45** we used the reaction conditions which resulted in the 62:38 selectivity observed in the competition experiments: equimolar amounts of peptide fragments **47** and **49** at a final 1 mM concentration in methanol mixed with 2 equivalents of DIPEA as non-nucleophilic base (Figure 4.5). Using this data, curves plotting the change in the concentration of peptide **45** throughout the first 6 hours of reaction were built (Figure 4.6).

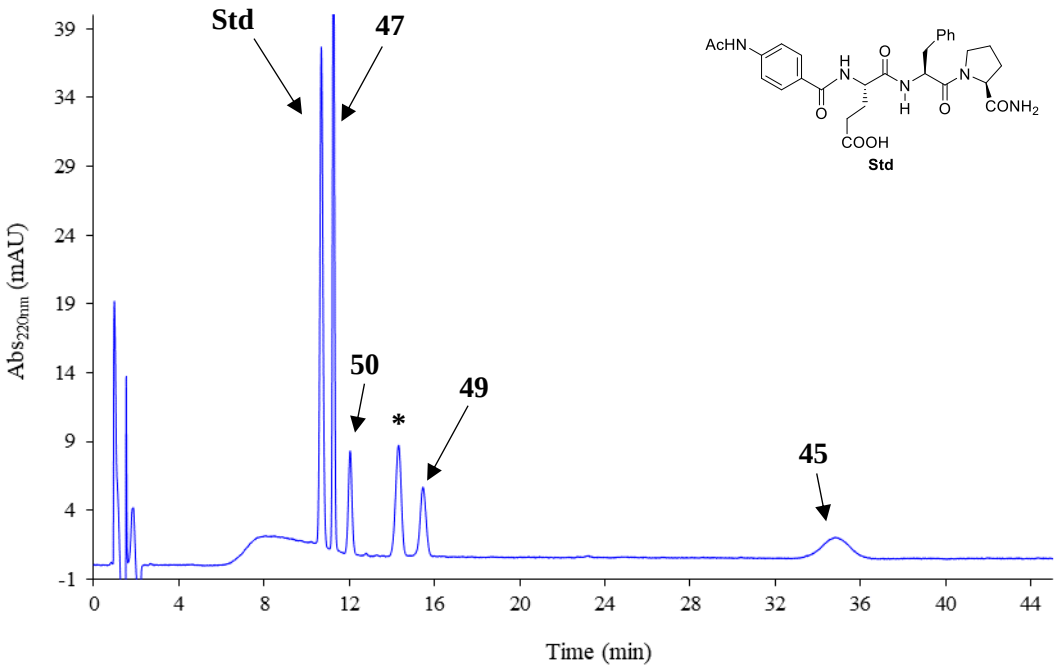


Figure 4.5: Representative HPLC traces of the ligation of fragments **47** and **49** in the presence of **45** seeds. In this case, a reaction at 1 mM peptide concentration in methanol with 2 equivalents of DIPEA, the reaction proceeded for 3 h with no externally added **45** (for full reaction and analytical conditions see section 7.5.2). * denotes a methyl ester derivative of **50** (free-carboxylic acid precursor) as detected by LC-MS. The structure embedded in the traces corresponds to peptide standard **Std**.

Firstly, the reaction in the absence of a **45** seed was analysed. The graph in Figure 4.6 shows that after an initial spike in the production of heptamer **45** over the first two hours, the rate of the coupling then slowed down. This was interpreted as a result of the side-reactions of electrophile **49**, where the succinimide ester was hydrolysed to the free acid **50** or converted into a methyl ester derivative slowing the bimolecular ligation rate. After approximately 6 hours, the yield of product **45** was close to 20%. Reactions containing a **45** seed were then trialled. Unfortunately, it was found that when template **45** was added to the coupling reaction at 97 µM it seemed to slightly hinder the coupling of fragments **47** and **49**, leading to lower conversions after 6 hours (~13%; Figure 4.6). Furthermore, increasing the initial catalyst loading to 240 µM produced no significant effect on the rate of the coupling (Figure 4.6). These were surprising results, since an interference of the template to the reactivity of any of the fragments was hard to visualise *a priori*.

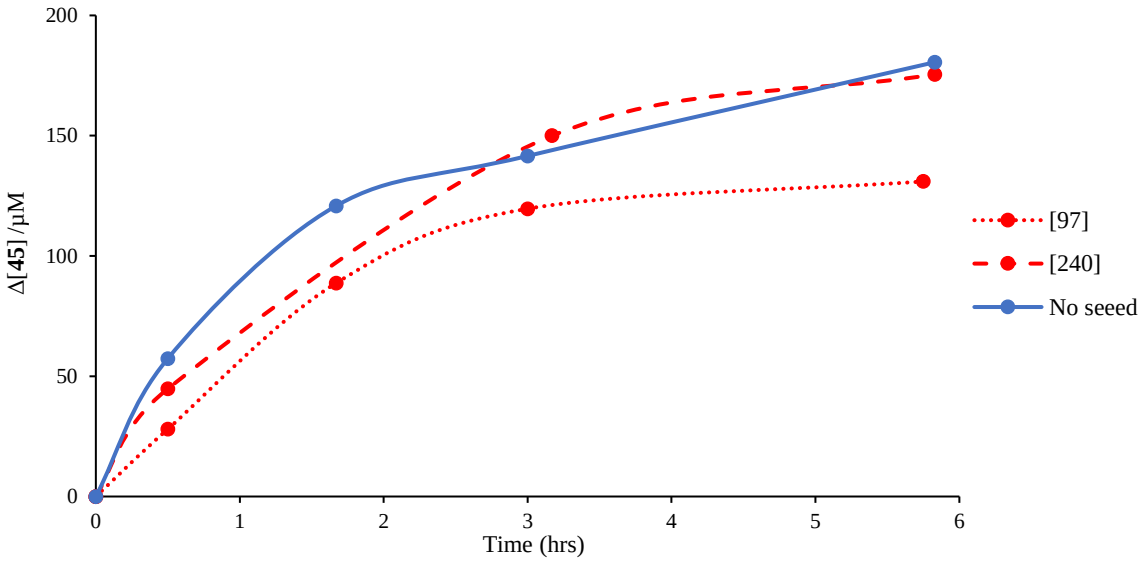


Figure 4.6: Production of 7-mer **45** over time by the coupling of fragments **47** and **49** in reactions seeded with different concentrations of **45**. Concentrations of the seeds are expressed as μM. The reactions were performed using 1 mM fragments in methanol with 2 equivalents of DIPEA (for full reaction and analytical conditions see section 7.5.2).

This initial set of ligations pointed to potential flaws in our standard coupling conditions which might be preventing the autocatalytic effect from being observed. First, we changed the solvent system for the coupling reactions; it was rationalised that a non-nucleophilic organic solvent could prevent the degradation of the electrophilic fragment, and at the same time, an aprotic organic solvent could improve the assembly of heptamer **45** into β-sheets, which would potentiate the templation effect. Therefore, we carried out couplings in a mixture of ACN/MeOH 9:1 using 1 mM fragments with 2 equivalents of DIPEA. Figure 4.7 shows that when using this solvent mixture, the coupling reaction proceeded more efficiently (25% conversion after 6 h) and presented a more linear trend. Secondly, we realised that the starting concentration of the fragments might have been too high. This could be making the ligation to proceed relatively fast or hinder the recognition process and the assembly of the peptide fragments with template **45**, which could be preventing us from observing the autocatalytic effect. For comparison, ligation reactions in literature self-replicating peptide systems were carried out with fragments at 200-300 μM concentrations with template loadings of 5-50%.[48,50] Thus, ligation of fragments **47** and **49** was attempted using the fragments at a 500 μM concentration (half of the original conditions). As expected, it was observed that at this concentration the ligation proceeded more slowly with a nearly linear trend (Figure 4.7; still, the reaction was relatively fast compared to literature ligations probing self-replication).

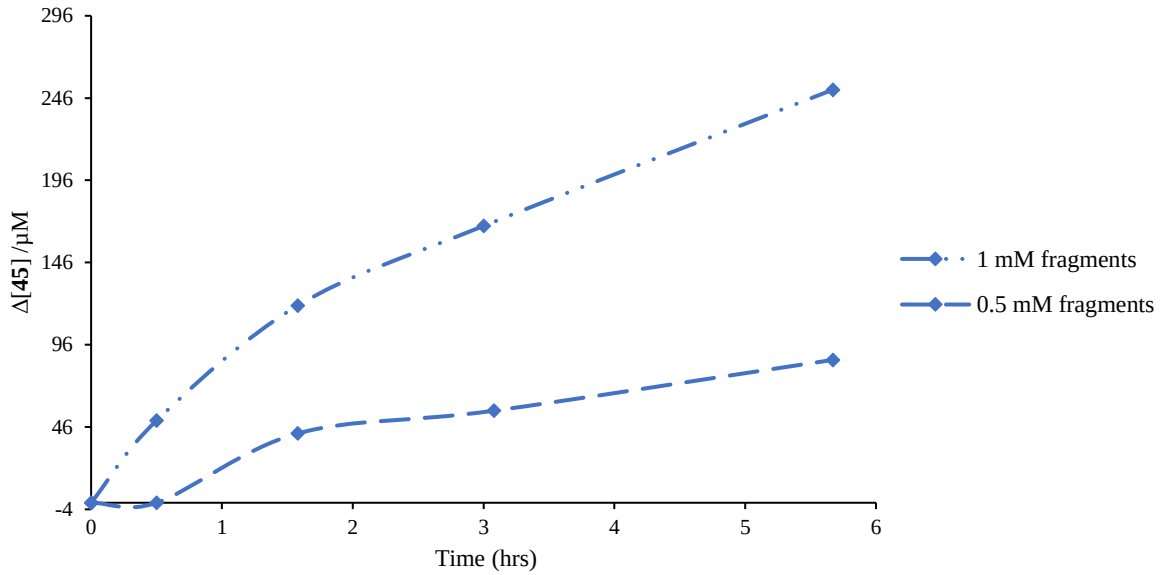


Figure 4.7: Production of 7-mer **45** over time by the coupling of fragments **47** and **49**. The reactions were performed in ACN/MeOH 9:1 with 2 equivalents of DIPEA (for full reaction and analytical conditions see section 7.5.2).

Under these new conditions, couplings of the peptide fragments in the presence of a **45** seed were probed. Unfortunately, it was found that adding template **45** seemed to have little to no effect on the ligation rate (Figure 4.8) and puzzlingly the largest seed (180 μM; Figure 4.8) seemed to be hindering the reaction slightly. It was hypothesis that this reactivity suppression could arise from an alternative unconsidered association between fragments and template which was unfavourable for the coupling reaction. These results were discouraging, moreover, it was also found that further decreasing the reactants concentration to 250 μM whilst using the same solvent system and adding a 105 μM template seed resulted in analogous results; no effect of seeding on the coupling rate.

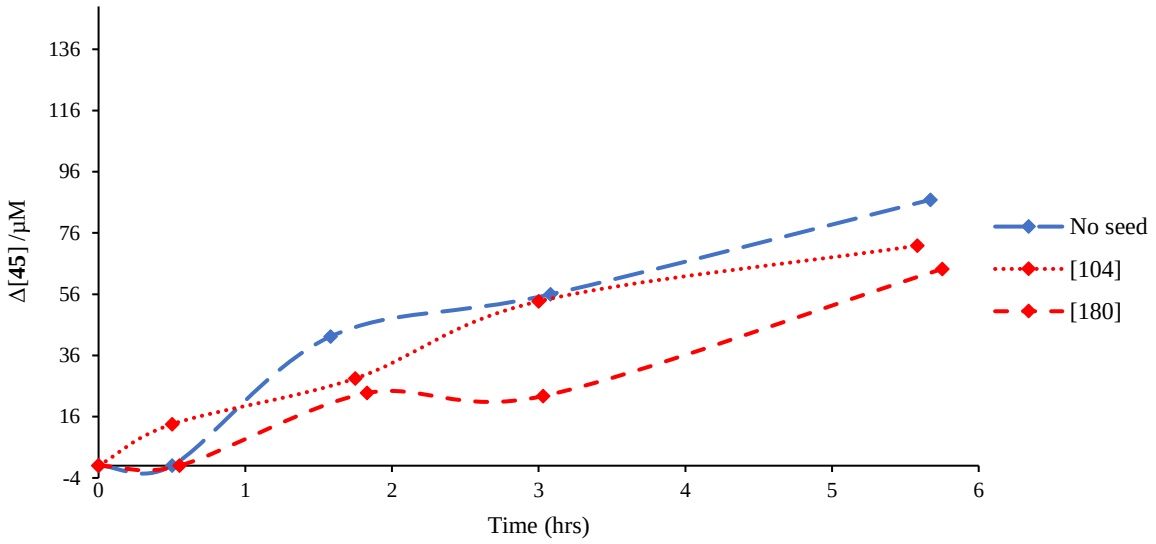


Figure 4.8: Production of 7-mer **45** over time by the coupling of fragments **47** and **49** in reactions seeded with different concentrations of **45**. Concentrations of the seeds are expressed as μM. The reactions were performed using 0.5 mM fragments in ACN/MeOH 9:1 with 2 equivalents of DIPEA (for full reaction and analytical conditions see section 7.5.2).

Finally, the ligation was tried in an alternative solvent: chloroform. Chloroform represented a logical choice over methanol, given our previous structural studies with the hairpin systems (chapter 3), which demonstrated the propensity of similar hydrophobic variants to adopt a β-sheet-like secondary structure in this solvent. Ligation of the fragments at 500 μM concentrations in the presence of 2 equivalents of DIPEA in chloroform proceeded slower than the analogue reaction in MeOH or ACN/MeOH 9:1 (Figure 4.9). Performing the coupling in chloroform with an added seed of template **45** at 110 μM presented little change to the efficiency of the coupling; a modest increase in the initial rate seemed to appear. However, this was disregarded, given that when the seed concentration was doubled (236 μM; Figure 4.9) such effect disappeared, and the reaction seemed practically unaltered by the seed or even slowed down during the first 2 hours. This behaviour does not correspond to a reaction proceeding autocatalytically.

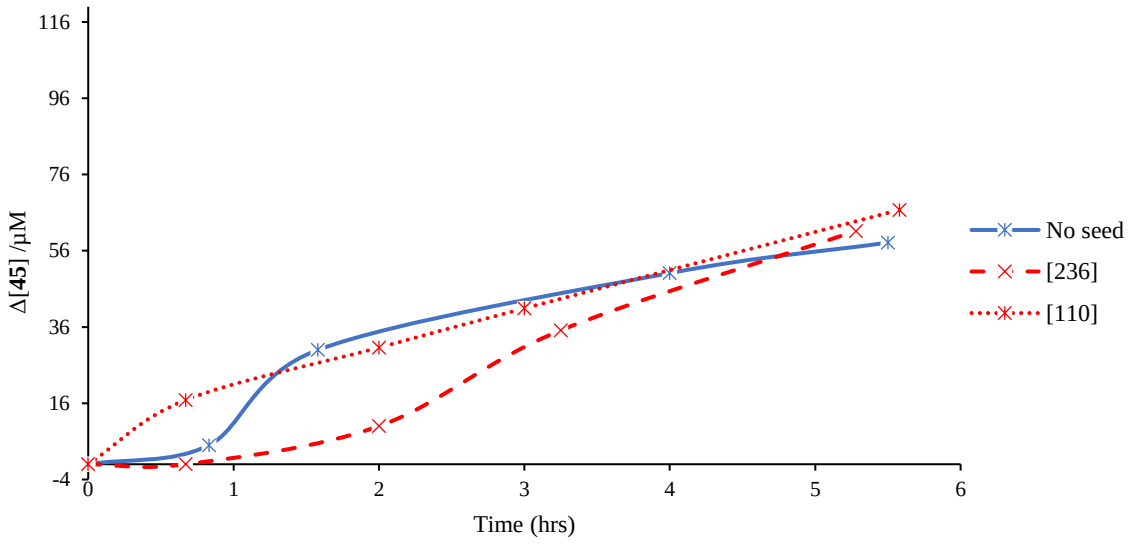


Figure 4.9: Production of 7-mer **45** over time by the coupling of fragments **47** and **49** in reactions seeded with different concentrations of **45**. Concentrations of the seeds are expressed as μM. The reactions were performed using 0.5 mM fragments in chloroform with 2 equivalents of DIPEA (for full reaction and analytical conditions see section 7.5.2).

4.7. A short structural study of the association of heptamer 45 to fragments 47 and 49

To help us rationalise these negative results, we studied the structure of the complex formed by template **45** with fragments **47** and **49**. A CD spectrum of the two fragments at a 250 μM concentration in ACN/MeOH 9:1 was recorded (in the absence of base; Figure 4.10). The spectrum showed two negative bands at 220 and 202 nm, which is indicative of the two fragments adopting a β-sheet-like structure to some extent, but also to the presence of some random coil conformation. A mixture containing template **45** at 100 μM and both fragments at 250 μM was then analysed (in the absence of base; Figure 4.10). It was noticed that the CD spectra adopted a pattern closer to that displayed by **45** alone (maximum at 196 nm and minimum at 220 nm; see Figure 4.3) characteristic of a β-sheet secondary structure. However, there was also a minimum at 205 nm, similar to that displayed by the two fragments alone, characteristic for a disordered secondary structure. These observations were used to explain the lack of autocatalysis in the ligation reactions, it was hypothesised that the β-sheet structure was not efficiently enforced by the template on the two fragments, therefore the recognition and templation was lost, which resulted in the absence of an autocatalytic effect.

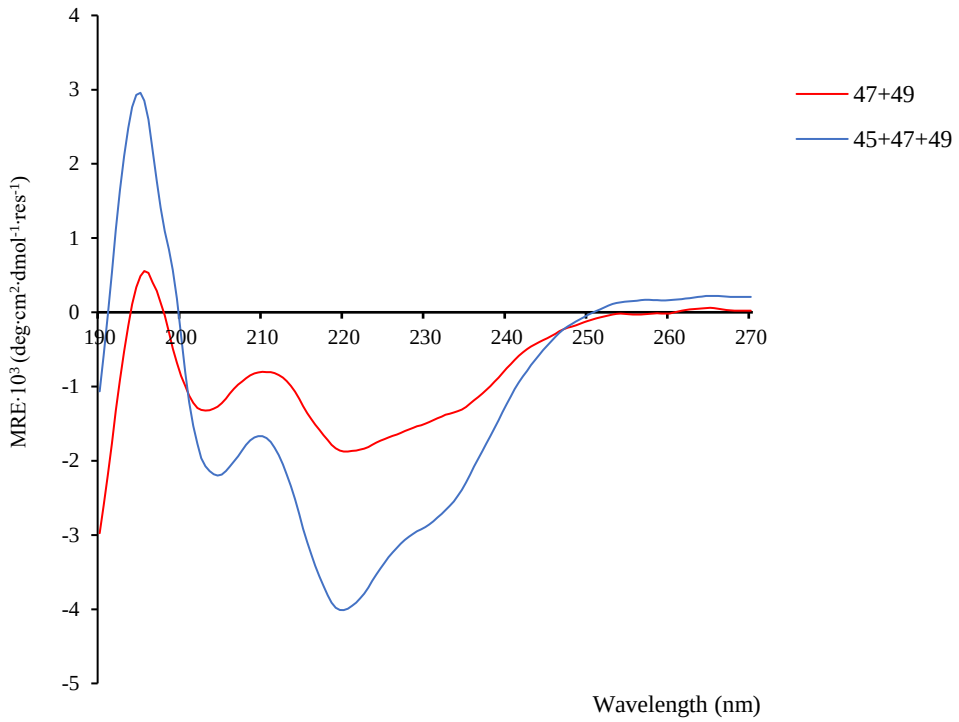


Figure 4.10: Mean residue ellipticity of a mixture of fragments **47** and **49** and a mixture of fragments **47**, **49** and heptamer **45** at 25 °C. Spectra are reported for ~250 µM peptide fragments and ~100 µM **45** in ACN/MeOH 9:1.

4.8. Conclusions and future work

A short hydrophobic PMC system based on two short templates (**45** and **46**) and three fragments (**47**-**49**) was designed. Using these peptides, a competition experiment where two nucleophilic fragments competed for a common electrophilic substrate resulted in an enrichment (62:38) of the self-complementary peptide **45** over the non-palindromic heptamer **46**. This selectivity pointed to a sequence-selective templation effect in the synthesis of the palindromic variant (**45**). To study this behaviour further, the rate of the formation of **45** was analysed. Unfortunately, so far, evidence of autocatalysis in the synthesis of palindrome **45** has not been obtained. Further screening of reaction conditions is currently under way, including the use of more diluted starting materials, low temperatures, less equivalents of base additive as well as probing the effect of pre-sonication of the template.

Alternatively, a short structural analysis suggested that a faulty pre-organization of the substrates and the template could possibly be causing the lack of autocatalysis. It was hypothesised that this faulty pre-organization originated from an inefficient promotion of a β-sheet secondary structure in the system, therefore, a longer template variant with a more defined β-sheet structure could offer a solution for this issue.

Chapter 5. Intramolecular template-directed formation of new covalent bonds by PMC

Unless stated otherwise, all results were obtained by the author of this thesis (Alberto Avila Castro)

5.1. Project aims

The selectivity observed in the competitive synthesis of PMC template **45** (see above) prompted us to design an alternative system to show the ability of PMC to transmit information. Given that it was hypothesised that a potential flaw of our organic-soluble peptide replicator came from a lack of efficient pre-organization of the constituting fragments by the short template, we were interested in a design that could enforce such pre-organization and sequence-selective recognition.

Therefore, we devised a system based on the β-hairpin motif where the sequence-selective recognition of α/β-residues would be used for the formation of new covalent bonds. A β-hairpin structure was divided into two fragments: "a shepherd's crook" containing the β-turn section (**51**) and a shorter fragment containing the N-terminus (**52**; Figure 5.1). We expected the β-turn to enforce the assembly of the two fragments into a β-hairpin structure, so that a short fragment (functionalised with an appropriate reactive terminus) bearing a matching α/β-sequence with respect to the strand in the shepherd's crook would be recognised by the crook and react rapidly with it to generate the β-hairpin bearing matching α/β-sequences. In contrast, an electrophilic fragment with a mismatching α/β-sequence (**53**) should not experience such templation. This would demonstrate the template-directed transmission of information by PMC.

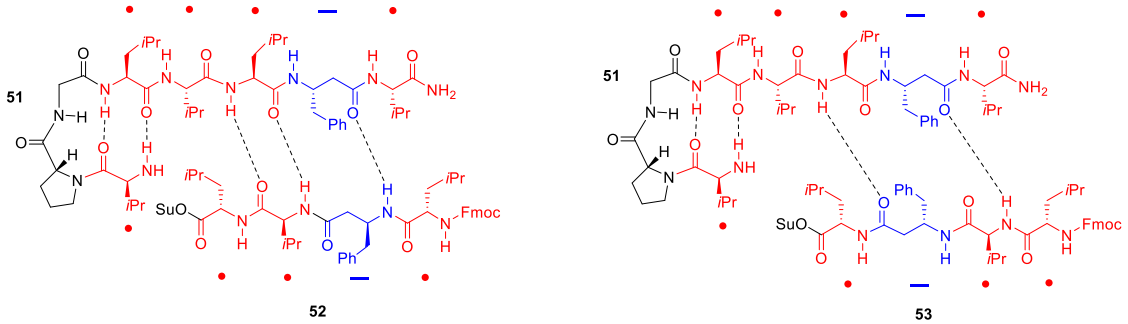


Figure 5.1: Structure of the **51**/**52** and **51**/**53** systems. Efficient recognition of the α/β-sequence in the **51**/**52** strands is contrasted to that of mismatching sequence in **51**/**53**. Su = succinimide.

5.2. Peptide design

We designed a system closely based on our hydrophobic linear templates and our β-hairpins. Two dodecamer β-hairpins **54** and **55** were built (Figure 5.2), the first bearing β-strands with matching α/β-sequences and the second bearing mismatching sequences. These molecules were based on decamers **35** and **38**, respectively. An extra pair of Leu residues were introduced, and the C-terminus was functionalised as an amide to avoid post-SPPS modification. The breakage of the amide bond at the Val8 position of these molecules would generate a shepherd's crook fragment with a nucleophilic free-NH_2 (**51**) and two shorter fragments (functionalised as *N*-hydroxysuccinimide esters). Fragment **52** bore a matching α/β-sequence and fragment **53** bore a mismatching one. The idea behind this three-fragment system was to carry out competition experiments analogous to those performed with the linear variants (see chapter 4); mixing the two electrophilic fragments **52** and **53** with a limiting amount of the nucleophilic fragment **51** should result in a selective production of match **54** over peptide **55**, due to the recognition and templation of the matching β-strands in the former. This templation should be absent in the production of mismatch **55** (Figure 5.2).

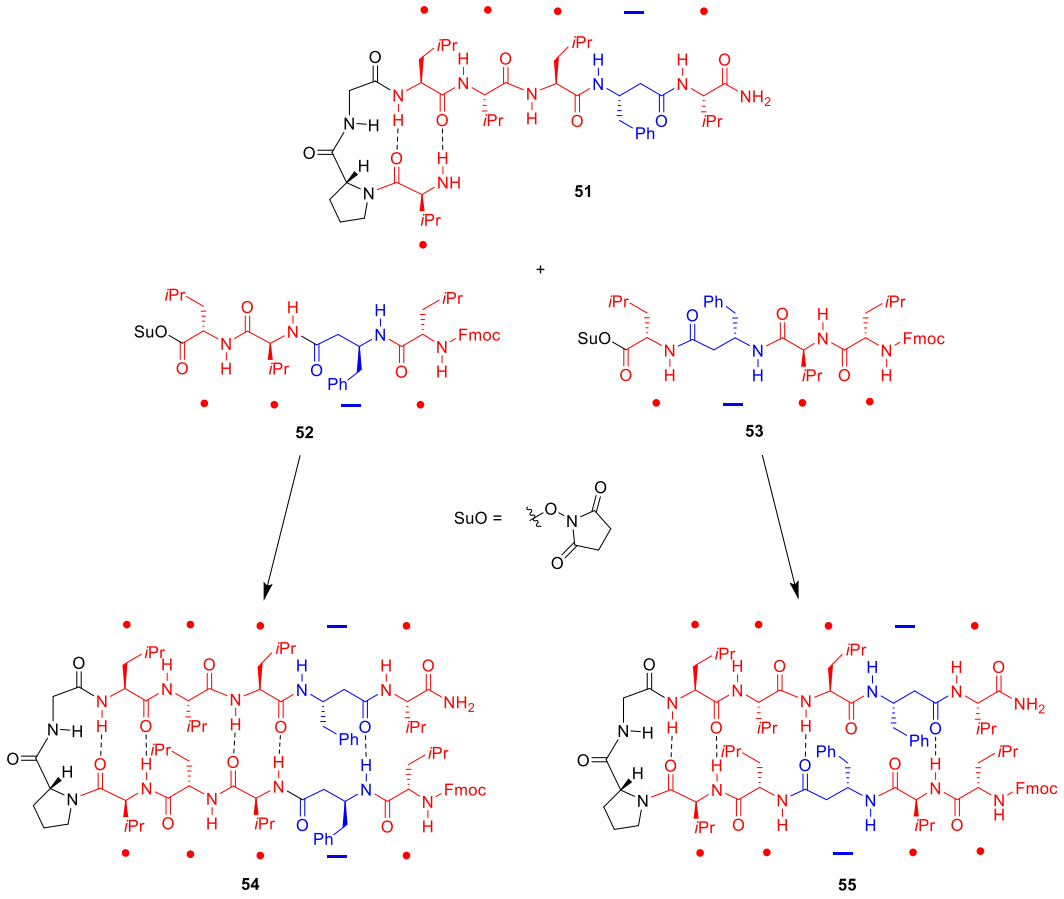
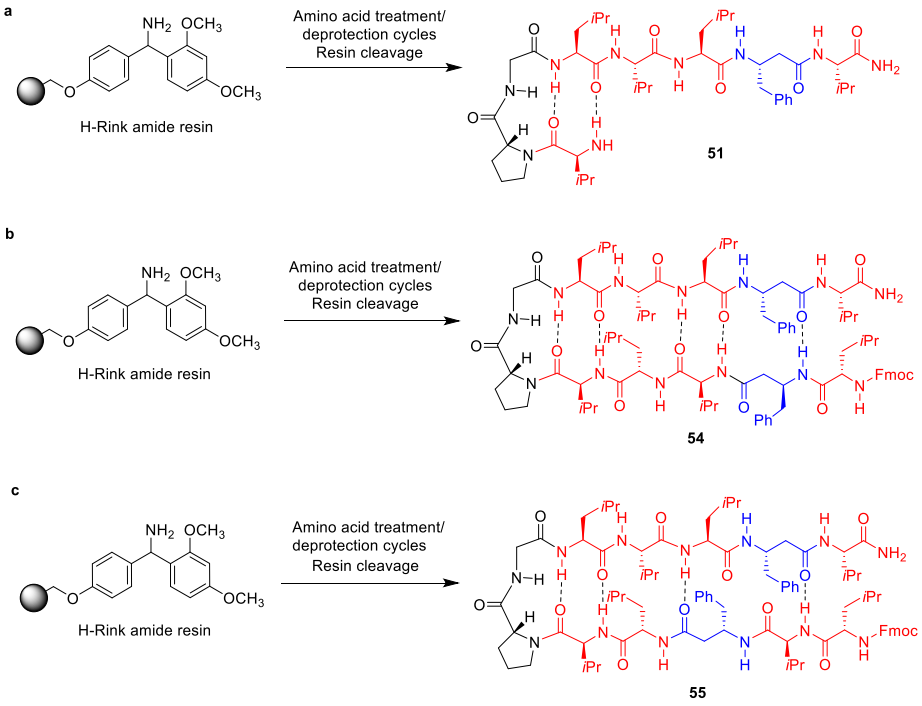


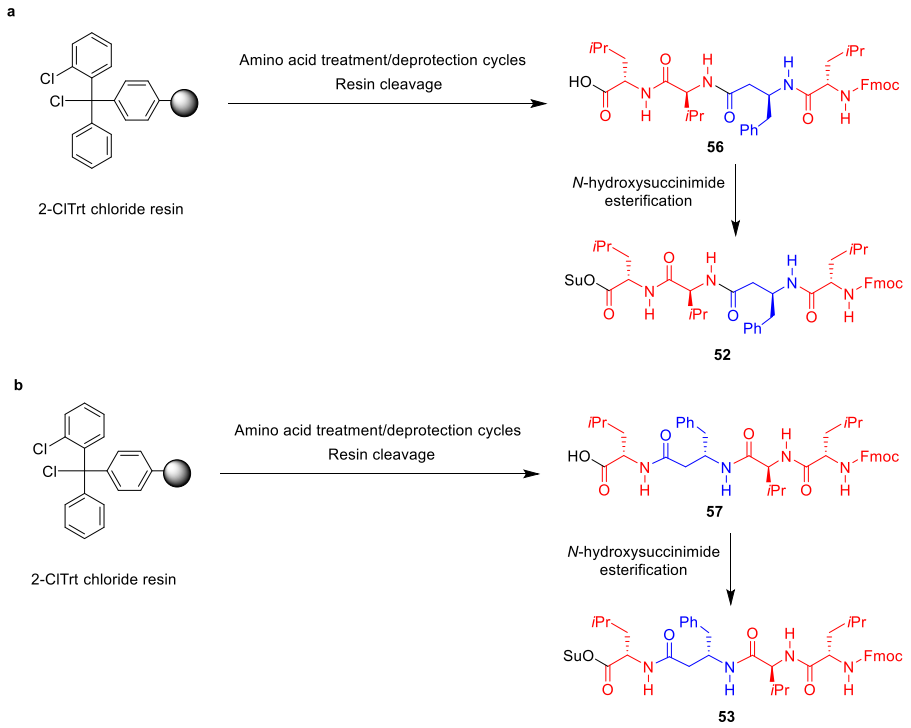
Figure 5.2: **Top**. Coupling of a mixture of the nucleophilic fragment **51** with activated esters **54** and **55** to generate dodecamers **54** and **55**. **Bottom**. The generated product **54** bears matching β-strands. In contrast, **55** is shown to adopt a less stable β-hairpin structure.

5.3. Peptide synthesis

The two β-hairpins and the three peptide fragments were synthesised using a standard Fmoc-based manual solid-phase peptide synthesis (SPPS) methodology (Schemes 5.1 and 5.2, see also sections 7.2.1 and 7.3.3). An H-Rink amide resin was used to obtain peptides **51**, **54** and **55**. The electrophilic fragments were synthesised as C-terminal carboxylic acids using a 2-chlorotrityl chloride resin (**56** and **57**). After cleavage, the free acids (**56** and **57** were used without further purification) were converted to the *N*-hydroxysuccinimide esters **52** and **53**.[107] The peptides were purified by preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA) or methanol/water). The identity and purity of the peptides was assessed by MALDI-TOF MS and analytical HPLC (see section 8.1.4).



Scheme 5.1: Overview of the synthesis of peptides **51** (**a**), **54** (**b**) and **55** (**c**; for full conditions see section 7.2.1).



Scheme 5.2: Overview of the synthesis of peptide fragments **52** (**a**) and **53** (**b**; for full conditions see sections 7.2.1 and 7.3.3).

5.4. Structural study of dodecamers 54 and 55

A brief study of the secondary structure adopted by the two dodecamers **54** and **55** was carried out by analysing their CD spectra (Figure 5.3). Spectra was recorded at ~100 μM in methanol. Match **54** presented a maximum close to 198 nm and a broad minimum at 224 nm. This represented a very similar pattern to that displayed by the matched β-hairpins analysed in chapter 3. On the other hand, the spectrum of mismatch **55** displayed two minima at 206 and 220 nm, an analogous spectrum to that of mismatch decamer **38**, corresponding to a peptide adopting both folded β-hairpin and random coil conformations (see section 3.3.6).

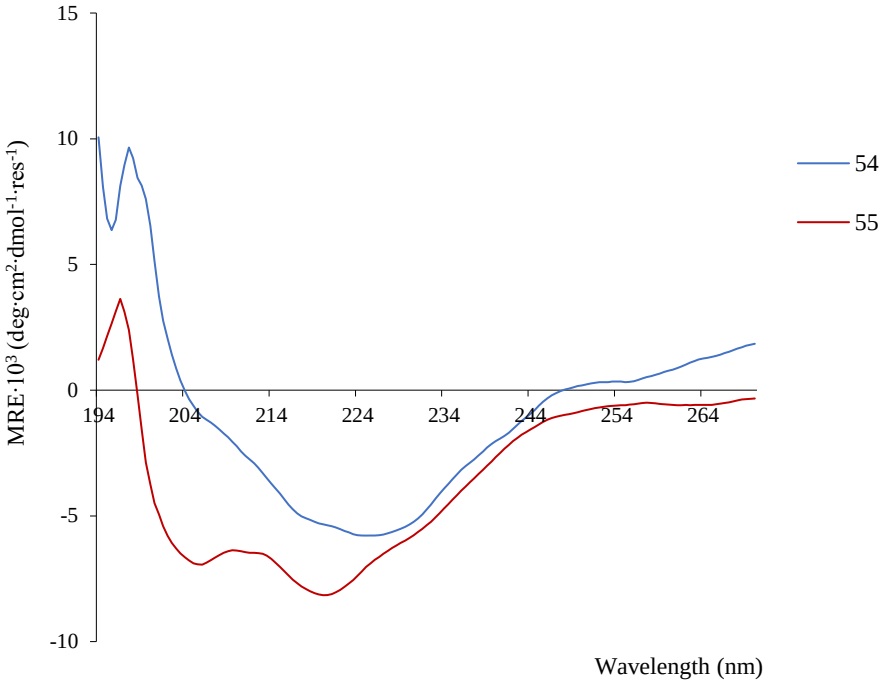


Figure 5.3: Mean residue ellipticity of peptides **54** and **55** at 25 °C. Spectra are reported for ~100 μM peptide samples in methanol.

5.5. Competition experiments

Once we had obtained evidence of the divergent structure adopted by dodecamers **54** and **55** due to their matched *vs* mismatched β-strands we set out to perform competing coupling reactions of their constituting fragments to probe the template-directing capabilities of our system (Figure 5.2). Initially, we envisioned to try the coupling of fragment **51** with electrophiles **52** and **53** using the same standard conditions developed for the linear hydrophobic system. However, it was found that both the β-hairpin products and their constituting fragments were less soluble than the shorter linear analogues. Therefore, the coupling reactions were carried out using equimolar amounts of the starting materials at 250 μM concentrations in the presence of 1.2 equivalents of DIPEA (see section 7.5.3). An analytical HPLC method was also developed to resolve the two coupling products **54** and **55**, consisting of a 40 minute-gradient from 70% to 100% methanol in water. Figure 5.4 shows representative HPLC traces of the coupling competition, peaks with t_R = 32.2 and 33.5 min correspond to **55** and **54**, respectively. The ratio of the products was determined by comparing the corresponding areas under the peak. The presence of some hydrolysis products of the *N*-hydroxysuccinimide esters was also indicated by LC-MS analysis.

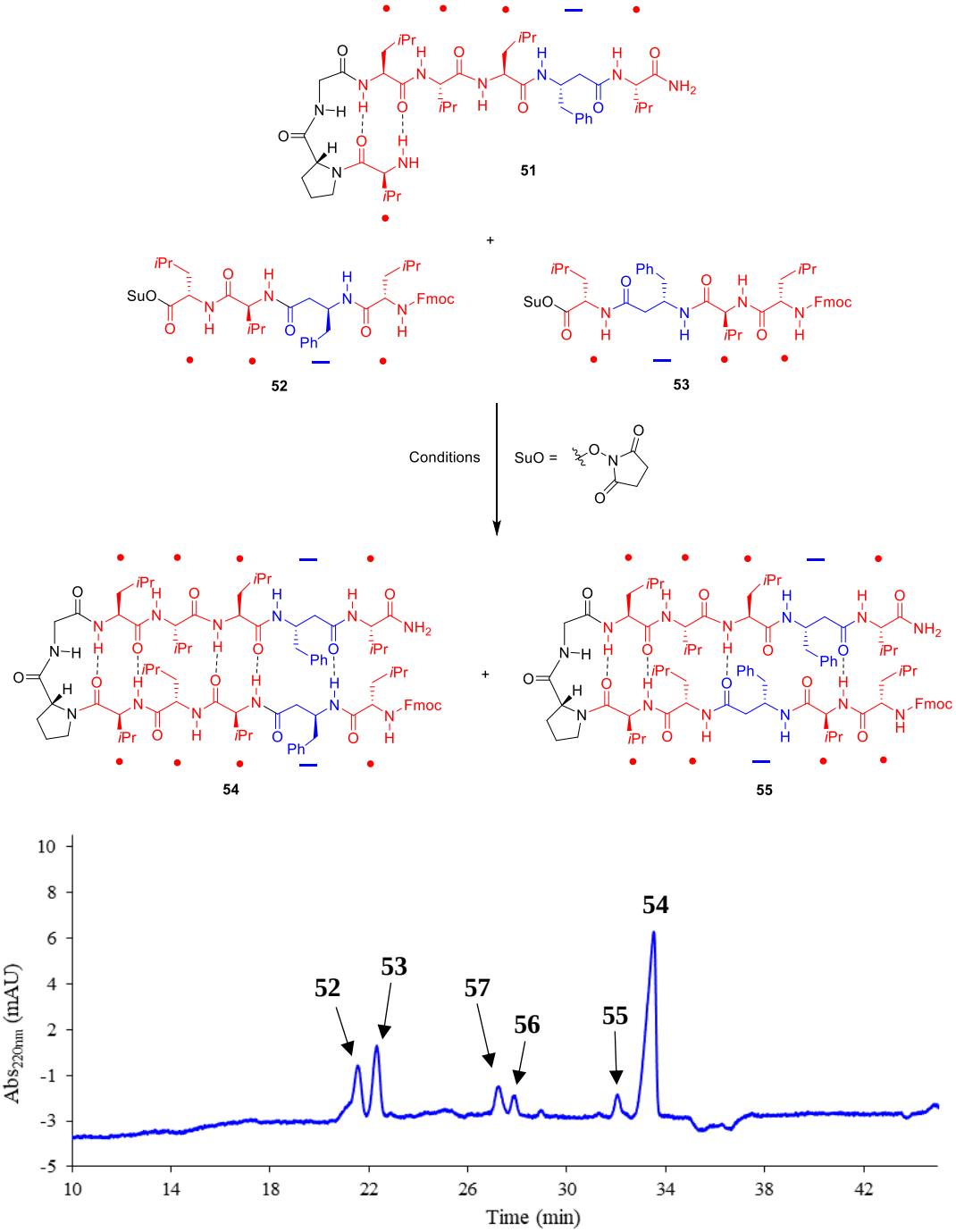


Figure 5.4: **Top**. Scheme of the coupling of fragments **52** and **53** (equimolar) with fragment **51**. **Bottom.** Representative HPLC traces of the competitive ligation of fragments **52** and **53** with **51**. In this case, a reaction in chloroform that had proceeded for 4 h (for full reaction and analytical conditions see section 7.5.3). Carboxylic acid derivatives **56** and **57** (Scheme 5.2) were detected by LC-MS.

Using the standard conditions, the ratio of products was evaluated after 4 hours of reaction time. Carrying out the couplings in methanol generated a ratio of **54** to **55** of 68:32, slightly better than 2:1. A range of solvents were then tested in an attempted to improve this selectivity (Table 5.1). Changing to a mixture of methanol/ACN 1:1 improved the selectivity to 80:20 and on using chloroform, the

selectivity reached nearly a 9:1 ratio. At this point, we limited the amount of nucleophilic fragment **51** to 0.9 equivalents, further increasing the selectivity to a 94 to 6 ratio (Table 5.1, entry 4). These were highly encouraging results which pointed to the template-assisted formation of match **54**, based on the matching information encoded in its α/β-residue sequence.

Table 5.1: Competitive ligation of electrophilic fragments **52** and **53** with fragment **51**.

Entry	Solvent	Ratio of 54 to 55
1	Methanol	68:32
2	Methanol/ACN 1:1	80:20
3	Chloroform	89:11
4[a]	Chloroform	94:6

[a]Using 0.9 equivalents of fragment **51**.

5.6. Conclusions and future work

A simple system based on three fragments **51**-**53** (a nucleophilic shepherd's hook and two shorter electrophiles) and two products **54** and **55** was designed to gain proof of the principle behind information transfer by PMC. To overcome the difficulties encountered with previously studied linear variants, we monitored the template-directed formation of new covalent bonds in structures forming intramolecular β-sheets. An experiment where the coupling of a set of electrophilic and nucleophilic fragments bearing matching α/β-sequences was carried out in the presence of a competing electrophile bearing a α/β-mismatching sequence to that of the nucleophile resulted in the selective generation of the product bearing matching strands over the mismatching one (94:6). This high selectivity provided strong evidence for a template-directed ligation, dependent on the α/β-residue sequence of the fragments.

Further analysis of this system is currently being performed to confirm these results. A control coupling reaction where the templating effect should be absent is underway. Such control is illustrated in Figure 5.5; we expect to observe an analogous reactivity of both *N*-hydroxysuccinimide esters with the Pro nucleophile. This should serve as a confirmation that the enrichment of match **54** over mismatch **55** originates from a template-assisted effect.

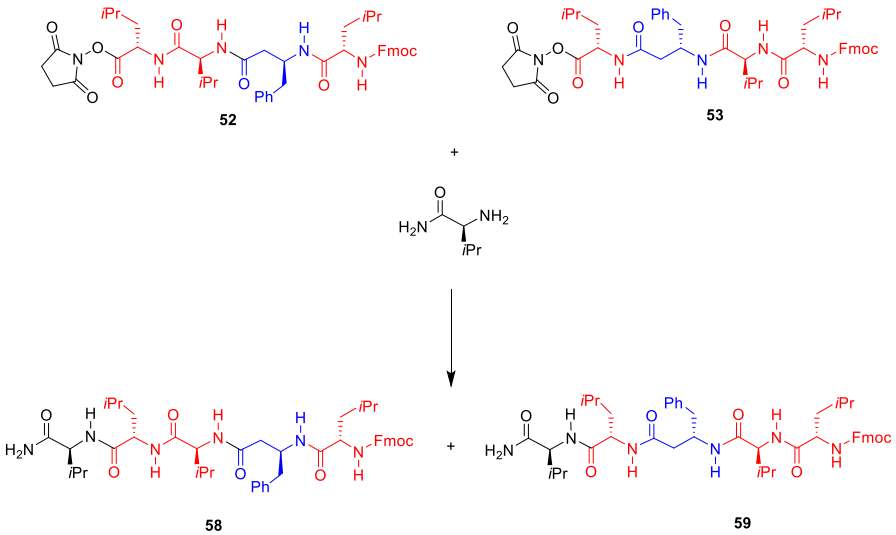


Figure 5.5: Control competition experiment where electrophilic fragments **52** and **53** are coupled with Val. Since no templation is expected from this system, no selectivity in the production of peptides **58** and **59** should be observed.

Chapter 6. Overall conclusions

Through the work described in this thesis, a series of short and simplified peptide systems have been designed to probe the information storage and transmission capabilities of α/β-peptide hybrid β-sheets: the peptide Morse code model, a model with potential prebiotic relevance. Our studies have provided the first evidence for both, the α/β-sequence-selective assembly of β-sheet structures and the template-directed formation of covalent bonds by these sequence-dependent structures.

6.1 PMC information storage

The system designed in chapter 3 comprised a series of short 10-mer α/β-peptide hybrids containing a ^{D}Pro-Gly segment (Figure 6.1). It was shown through NMR and CD analyses that molecules with matching α/β-residues in their β-strands would fold into stable β-hairpins in organic media, due to the formation of an intramolecular β-sheet. On the other hand, there was structural evidence for the absence of such stable intramolecular β-sheets in molecules bearing strands with mismatching α/β-residues. This behaviour extended to larger molecules with longer β-strands, however further studies of such molecules and possibly design modifications would be needed to confirm our conclusions on those larger assemblies. Work is currently ongoing to synthesise β^{3}hVal-based dodecamers.

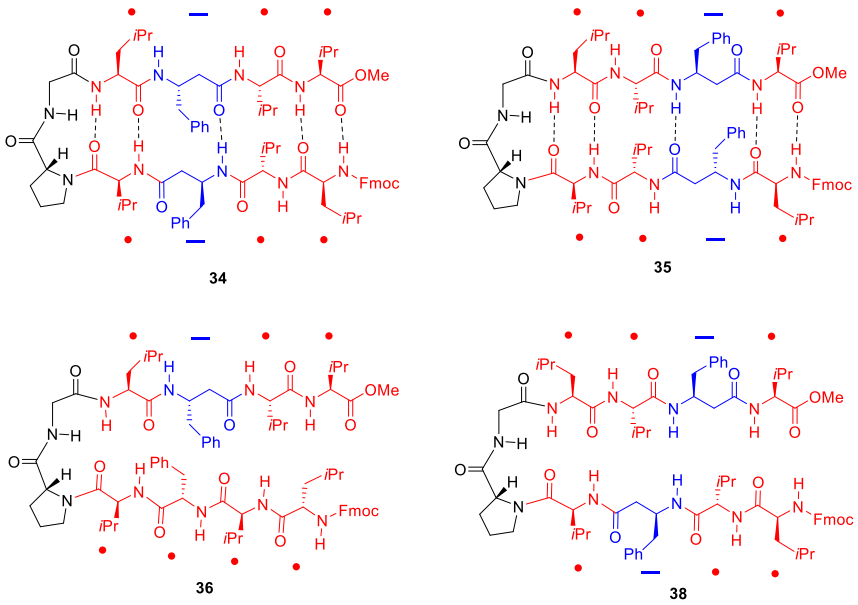


Figure 6.1: **Top**. α/β-peptide hybrid 10-mers bearing matching α/β-strands. **Bottom**. α/β-peptide hybrid 10-mers with mismatching strands.

Moreover, in chapter 4 we showed the difference in the CD spectra of two short linear hydrophobic α/β-peptide hybrids (Figure 6.2); heptamer **45** bearing a palindromic α/β-sequence which could form antiparallel β-sheets with itself and **46**, an heptamer with a non-self-complementary sequence of α/β-residues for which formation of β-sheets should be discouraged. Put in the context of our decamer β-hairpin system, the CD analysis pointed to a difference in the propensity of these peptides to form β-sheets. Longer designs based on these heptamers are currently underway.

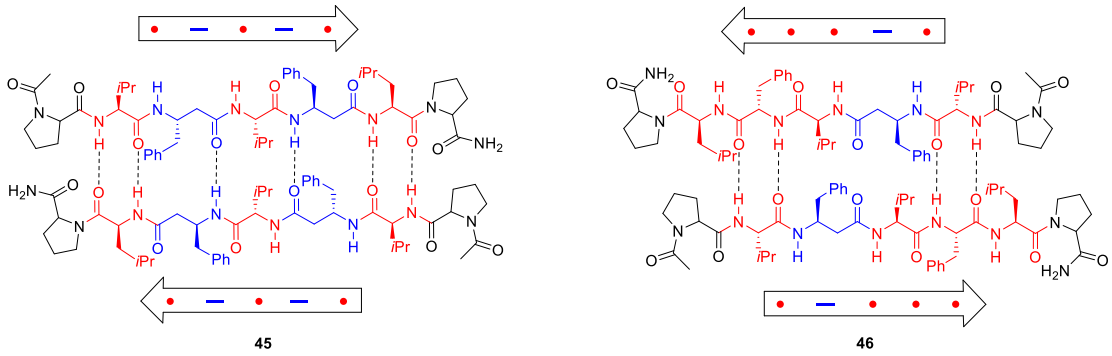


Figure 6.2: Structure of α/β-peptide hybrid 7-mers **45** (bearing a palindromic sequence) and **46** (bearing a non-palindromic sequence).

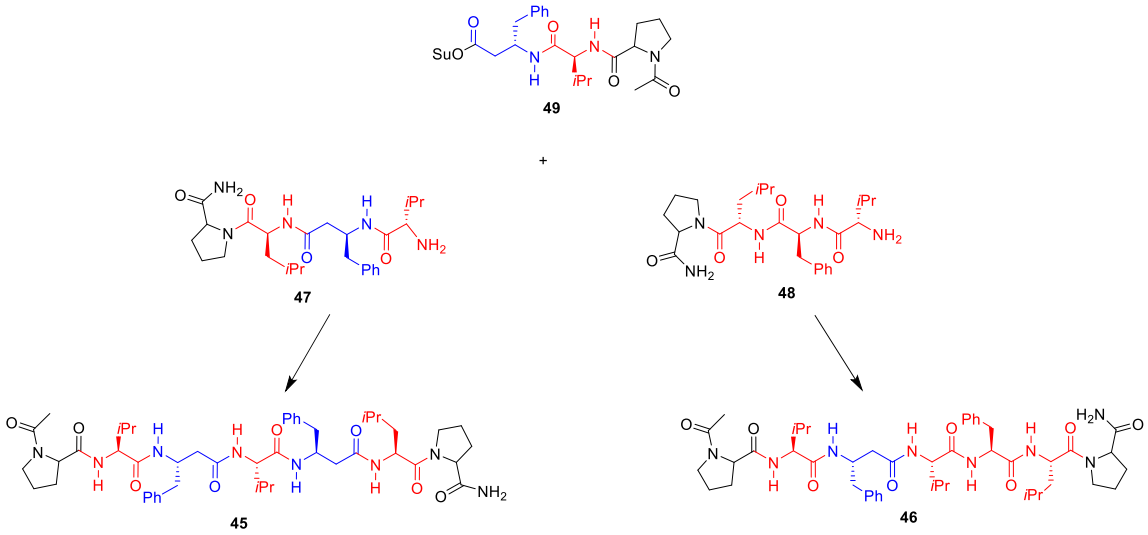
In summary, the results obtained from the intramolecular and short intermolecular systems have established that α/β-peptides have the potential for information storage, through the sequence-selective formation of β-sheets, the structural basis of PMC.

6.2. PMC information transmission

Our most ambitious studies focused on evaluating sequence-selective self-replication in α/β-peptide hybrids. We first designed a water-soluble amphiphilic PMC variant of a literature peptide replicator. However, autocatalysis in this system could not be evaluated due to various experimental difficulties. Furthermore, the reported autocatalysis of the model peptide replicator proved hard to reproduce which prompted us to design an alternative α/β-peptide system.

Based on the structural studies carried out with hydrophobic peptide variants, we built a system based on the two heptamers **45** and **46** (Figure 6.2). A competing ligation experiment where two constituting fragments of the heptamers competed for a third fragment common to both heptamers showed a selectivity (62:38) for the production of the self-complementary molecule (**45**; Scheme 6.1). This provided initial evidence for a template-directed mechanism in the coupling of the fragments. However, further studies on the kinetics of the self-replication of **45** have not yet shown signature autocatalytic

behaviour in the production of this molecule. Further conditions for this reaction are currently being screened in search of that signature behaviour.



Scheme 6.1: Coupling of a mixture of the nucleophilic fragments **47** and **48** with activated ester **49** to generate heptamers **45** and **46**. Su = succinimide.

Taking inspiration from the β-hairpin system displayed in Figure 6.1 which had shown selectivity for the intramolecular formation of antiparallel β-sheets, we devised a system to probe α/β-sequence-selective templation of new covalent bonds (Figure 6.3). By performing competing coupling experiments analogous to those described above, we were able to show that peptide fragments bearing matching α/β-sequences (**51**/**52**) were selectively coupled at the expense of a mismatching set (**51**/**53**) in a significative ratio of 94:6. Controls for this experiment are currently underway.

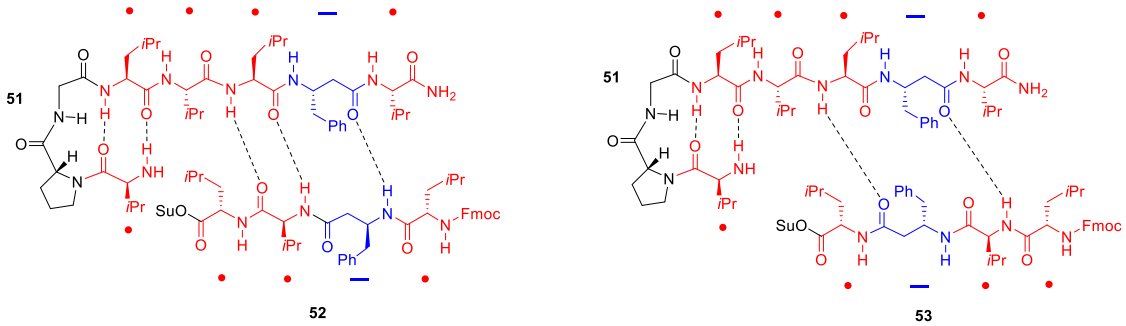


Figure 6.3: Structure of the **51**/**52** and **51**/**53** systems. Efficient templation in the **51**/**52** system is contrasted to that of mismatching **51**/**53**. Su = succinimide.

In summary, the results obtained from the competing ligations in the short linear and intramolecular systems have provided preliminary evidence for the ability of α/β-peptide hybrids to serve as templates for the formation of new covalent bonds, the basis for PMC information transmission.

Chapter 7. Materials and methods

7.1. General considerations

Peptide grade *N*,*N*-dimethylformamide (DMF) was purchased from Life Technologies (Paisley, UK). 3-[Bis(dimethylamino)methyliumyl]-3H-benzotriazol-1-oxide hexafluorophosphate (HBTU) and ***N***-(9-Fluorenylmethoxycarbonyl)-L-proline (Fmoc-Pro-OH) were purchased from Merck Chemicals (Beeston, UK). Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBop) and (S)-3-(Fmoc-amino)-4-phenylbutyric acid (Fmoc-β-Homophe-OH) were purchased from Fischer Scientific (Loughborough, UK). The rest of Fmoc protected amino acids, resins, solvents and other chemicals were acquired from Sigma Aldrich (Gillingham, UK), and used without further purification.

7.2. Peptide synthesis, purification and characterization

7.2.1. Solid-phase peptide synthesis

All peptides were obtained by standard Fmoc solid-phase synthesis on a 0.1 mmol scale. For each amino acid coupling step the resin was treated with 3 equivalents of the respective amino acid, 3 equivalents of HBTU (0.1 g, 0.3 mmol) as activating agent and 4 equivalents of DIPEA (0.1 mL, 0.4 mmol) in DMF (2 mL), shaking the mixture for 1 to 2 hours. Following the amino acid coupling, the resin was washed with DMF (3 × 15 mL). Fmoc deprotection was carried out using 20% (v/v) piperidine in DMF (3 mL) shaking the mixture for 30 minutes.[108] These coupling/washing/deprotection cycles were repeated until the desired sequence was obtained. To produce peptides as the C-terminal amides, a H-Rink-amide ChemMatrix resin was used. To produce peptides as the C-terminal carboxylic acids, 2-ClTrt chloride resin[vii] or 2-ClTrt resins with the first amino acid in the sequence already installed were used.

With regard to the N-terminus, different procedures were followed to generate the required functionalities, all with the peptide still attached to the solid support: to obtain N-terminally acetylated derivatives, the peptide was treated with 10 equivalents of AcOH (0.06 mL, 1.0 mmol), 10 equivalents of HBTU (0.38 g, 1.0 mmol) and 12 equivalents of DIPEA (0.2 mL, 1.2 mmol) in DMF (2 mL) for 2 hours. In the case of peptide **43**, the N-terminal $(EG)_3$ carbamate was produced by treating the peptide with 20 equivalents of PEG_3-succinimidyl carbonate[viii] (0.7 mL, 2 mmol) and 22 equivalents of DIPEA

[vii] The first residue in the sequence (β^3hPhe) was installed by treating the 2-chlorotrityl resin with 3 equivalents of the amino acid and 4 equivalents of DIPEA in DCM for 2 hours.

[viii] This compound was synthesised by treating triethylene glycol monomethyl ether (0.96 mL, 6 mmol) with 1.5 equivalents of *N,N′*-disuccinimidyl carbonate (2.3 g, 9 mmol) and 1.5 equivalents of DIPEA (1.6 mL, 9 mmol) in ACN overnight. The reaction mixture was then filtered, and the solvent was removed under vacuum yielding a

(0.36 mL, 2.2 mmol) in DMF (2 mL) for 4 hours. To produce ABA-capped derivatives, the peptide was treated with 10 equivalents of ABA (0.18 g, 1.0 mmol), 10 equivalents of HBTU (0.38 g, 1.0 mmol) and 12 equivalents of DIPEA (0.2 mL, 1.2 mmol) in DMF (2 mL) for 2 hours. To obtain the Fmoc protected derivatives, the last deprotection step was omitted, conversely, to obtain a N-terminal free amine the last residue was deprotected prior to cleavage from the resin.

Cleavage from the solid support was effected with a mixture of AcOH/TFE/DCM 1:1:8 (v/v/v) for 2-ClTrt resins. Cleavage from the H-Rink-amide resin and side chain deprotection were carried out using a mixture of TFA/H_2O/TIPS 9.5:2.5:2.5 (v/v/v). Following cleavage and/or side chain deprotection, volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding 45 mL of diethyl ether, the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried to yield a white solid.

7.2.2. Peptide purification

Crude peptides were purified by reverse-phase high performance liquid chromatography (HPLC) using a Waters 600 Controller with a Waters 2998 Photodiode Array Detector on a XSELECT CSH Prep C18 5µm OBD (10x250 mm) column at a 5.0 mL/min pump rate for semi-preparative scale and on a XSELECT CSH Prep C18 5µm OBD (19x250 mm) column at a 15.0 mL/min pump rate for preparative scale. Gradients of methanol in water or acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) were typically used. Absorbance was monitored at 220 and 270 nm. Pure fractions were identified by analytical HPLC and MALDI mass spectrometry (*vide infra*) and were combined and lyophilised.

7.2.3. Analytical HPLC

Peptide purity was confirmed by analytical HPLC using two systems: a Gilson 321 pump equipped with a Gilson 156 UV/Vis Detector and a Gilson 402 syringe pump on a Jupiter Phenomenex C4 5µm (4.6 × 150 mm) column at a 1.0 mL/min pump rate with step gradients of acetonitrile in 0.1 M NH_4HCO_3 (aq). Or, a Waters 600 Controller with a Waters 2998 Photodiode Array Detector on a XSELECT CSH C18 5µm (4.6 × 150 mm) column at a 1.0 mL/min pump rate, using step gradients of methanol in water or acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA). Absorbance was monitored at 220 and 270 nm. Representative analytical HPLC traces for each peptide are shown in section 8.1.

dense transparent liquid (0.9 g, 3.0 mmol, 50% crude yield). The product was used without further purification. Reference: Hirotsune, S. Preparation of lissencephaly therapeutic agent WO 2012074120, June 07, 2012.

7.2.4. Analytical LC-MS

Samples were analysed using a Waters autopurification system comprising a SFO, 2767 autosampler and 2545 pump with a XSELECT CSH C18 3.5µm (4.6 × 100 mm) column at a 1.0 mL/min pump rate using step gradients of acetonitrile (0.1% v/v FA) in water (0.1% v/v FA). Detection was carried out using a Waters 2998 diode array detector monitoring between 210 and 600 nm and a Waters SQD2 ESI mass spectrometer detecting in positive mode between 150 and 1800 m/z.

7.2.5. Mass spectrometry

Matrix-assisted laser/desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and Matrix-assisted laser/desorption ionization tandem time-of-flight mass spectrometry (MALDI-TOF-TOF) for peptide sequencing were performed on a Bruker Daltonics UltrafleXtreme operating in both positive and negative reflector modes, using α-Cyano-4-hydroxycinnamic acid (CHCA) as ionization matrix (10 mg CHCA in 1 ml of a H_2O/ACN/Formic acid 49.9:49.9:0.2 v/v solution). Calculated masses quoted are for the monoisotopic singly charged species. Measured mass is to 0.1% accuracy. Representative mass spectra for each peptide are shown in section 8.1. Nanospray ionization mass spectrometry (NSI-MS) and Electrospray ionisation high resolution mass spectrometry (ESI-HRMS) were acquired through the University of Bristol mass spectrometry service.

7.2.6. Nuclear magnetic resonance spectroscopy

^{1}H and ^{13}C NMR spectra were recorded on a Varian VNMR 500 MHz, Bruker Advance III HD Cryo 500 MHz and a Varian VNMRS (equipped with a ^{1}H observe cryogenically cooled probe) 600 MHz spectrometers. All spectra were obtained at ambient temperature unless stated otherwise. All ^{1}H and ^{13}C NMR chemical shifts are reported relative to the ^{1}H (residual) and ^{13}C chemical shifts of the solvent as standard. The peptides were assigned using standard 2D homonuclear and heteronuclear spectra: TOCSY, NOESY, ROESY, COSY, HSQC and HMBC. Unless otherwise stated, 200 ms mixing times were used for NOESY and ROESY collecting 64 scans. Characterization data for the analysed peptides is summarised in section 7.6 and spectra for selected peptide is shown in section 8.3.

7.2.7. Variable temperature NMR experiments

Variable temperature experiments were carried out in a Varian VNMR 500 MHz spectrometer. ^{1}H NMR and TOCSY NMR spectra was obtained at each monitored temperature to follow the NH chemical shifts. Spectra was recorded at 4 different temperatures, ranging from 25 to 40 °C at 5 degree intervals. The sample was allowed to equilibrate at the desired for 1 minute before recording the spectra. After the experiment, the sample was cooled back down to 25 °C and the spectra obtained were not different to those initially recorded at this temperature. Partial spectra at the recorded temperatures are shown in section 8.4.

7.2.8. Circular dichroism spectroscopy

CD spectra were collected using a JASCO 810 spectropolarimeter, the measurements were carried out in 1 mm pathlength cuvettes. The spectra were baseline corrected and recorded as the average of eight scans from 270 to 190 nm at room temperature using a scanning speed of 100 nm/min, a 0.5 nm interval, a bandwidth of 1 nm, and a 1 s response time. Data was solvent subtracted and then spectra were converted from ellipticities (mdeg) to mean residue ellipticities (MRE deg·cm^{2}·dmol^{-1}·res^{-1}) by normalising for peptide concentration, number of peptide bonds and the cuvette pathlength using the equation:

(7.1) $$[\text{MRE}]\ (\text{deg}\cdot\text{cm}^{2}\cdot\text{dmol}^{-1}\cdot\text{res}^{-1}) = \frac{\theta * 100}{c * l * b}$$

where the variable θ is the measured difference in absorbed circularly polarised light (mdeg), *c* is the concentration of the sample (mol·cm^{3}), *l* is the sample pathlength of the cuvette (cm) and *b* is the number of amide bonds in the peptide. Representative spectra for the studied peptides are shown in section 8.2. The high-tension voltage and absorbance spectra were also recorded. The final concentration of each peptide solution was determined by absorbance of Fmoc at 267 nm[109] using the Beer-Lambert Law equation:

(7.2) $$[\text{A}] = \varepsilon * c * l$$

Where A is absorbance, is the extinction coefficient (mol^{-1}·dm^{3}·cm^{-1}), *c* is the concentration (mol·dm^{-3}) and *l* is the pathlength (cm). In the case of peptides that did not contain the Fmoc group the concentration was determined using a literature protocol based on the absorbance at 214 nm.[110]

7.2.9. Dynamic light scattering

Samples were prepared dissolving lyophilised peptides in ammonium carbonate buffer (pH 7) or Millipore water to yield the desired concentrations. The samples were filtered through a 0.45 μm membrane filter into an optical glass cuvette (10.0 mm path length). The measurements were performed on a Malvern Instruments Zetasizer Nano S using a 5 mW He-Ne laser (633 nm) at room temperature. The correlation function was acquired in real time and analysed with a function capable of modelling multiple exponentials. This process enabled the diffusion coefficients for the component particles to be extracted, and these were subsequently expressed as effective hydrodynamic radius, by volume, using the Stokes-Einstein relationship.

7.3. Post-SPPS modifications

7.3.1. Synthesis of peptide thioesters

22

Crude peptide **21a** (70 mg, 78.6 μmol) and benzyl mercaptan (29 mg, 0.2 mmol) were dissolved in DMF (2 mL) and stirred for 5 minutes. A solution of PyBop (123 mg, 0.2 mmol) and DIPEA (40 μL, 0.2 mmol) in DMF (1 mL) was added and the mixture was stirred for 1.5 hours. The reaction mixture was then precipitated by adding distilled water (45 mL), the solid was recovered by centrifugation and treated with 10 mL of a TFA/H_2O/TIPS 9.5:2.5:2.5 (v/v/v) mixture for 2 hours. Volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding 45 mL of diethyl ether, the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried. The peptide was purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield **22** (46 mg, 48.7 μmol, 62%) as a white solid.

23

Crude peptide **21a** (10 mg, 11.2 μmol) and sodium 2-mercaptoethanesulfoate (MESNA, 37 mg, 0.22 mmol) were dissolved in DMF (2 mL) at -10 °C and stirred for 5 minutes. A solution of PyBop (18 mg, 33.7 μmol) and DIPEA (6 μL, 33.7 μmol) in DMF (1 mL) was added and the mixture was stirred for 1.5 hours. The solvent was removed at reduced pressure yielding a white paste which was re-dissolved in H_2O/ ACN 1:1 (30 mL, v/v) and freeze-dried overnight. The resulting solid was treated with of a TFA/H_2O/TIPS 9.5:2.5:2.5 (5 mL, v/v/v) mixture for 2 hours. Volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding diethyl ether (45 mL), the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried. The

peptide was purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield **23** (4 mg, 4.3 µmol, 38%) as a white solid.

23'

Crude peptide **21a** (50 mg, 56.2 µmol) was pre-activated with PyBop (0.15 g, 0.28 mmol) and DIPEA (50 µL, 0.28 mmol) for 1.5 hours in DMF (4 mL) and subsequently treated with MESNA (50 mg, 0.28 mmol) for 1.5 hours at room temperature. The solvent was removed at reduced pressure yielding a yellow paste which was re-dissolved in H_2O/ ACN 1:1 (30 ml, v/v) and freeze-dried overnight. The resulting solid was treated with TFA/H_2O/TIPS 9.5:2.5:2.5 (10 mL, v/v/v) mixture for 2 hours. Volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding diethyl ether (45 mL), the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried. The **23-23'** epimers mixture was isolated by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield peptide **23'** (D-Ala, 5 mg, 5.6 µmol, 10%) as a white solid.

24

Crude peptide **21b** (20 mg, 22.8 µmol) and MESNA (10 mg, 68.5 µmol) were dissolved in DMF (2 mL) at room temperature and stirred for 5 minutes. A solution of PyBop (36 mg, 68.5 µmol) and DIPEA (12 µL, 68.5 µmol) in DMF (0.5 mL) was added and the mixture was stirred for 1.5 hours. H_2O/ ACN 1:1 (30 mL, v/v) was added to the reaction mixture and it was freeze-dried overnight. The resulting solid was treated with a TFA/H_2O/TIPS 9.5:2.5:2.5 (5 mL, v/v/v) mixture for 2 hours. Volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding diethyl ether(45 mL) , the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried. The peptide was purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield **24** (8 mg, 8.2 µmol, 36%) as a white solid.

25

Peptide thioester **22** (2.3 mg, 4.8 µmol) was dissolved in 6 M GdmCl/ 0.2 M phosphate buffer at pH = 7.1 (0.25 mL) and treated with MPAA (25.2 mg, 0.15 mmol) for 1 hour (final pH adjusted to 7.0). The peptide was precipitated using diethyl ether (30 mL), re-dissolved in H_2O/ACN 1:1 (20 mL, v/v) and freeze-dried overnight. The peptide was purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield **25** (2 mg, 4.1 µmol, 85%) as a white solid.

30

Crude peptide **29** (50 mg, 54.7 µmol) and MESNA (20 mg, 1.1 mmol) were dissolved in DMF (2 mL) at -10 °C and stirred for 5 minutes. A solution of PyBop (85 mg, 0.17 mmol) and DIPEA (30 µL, 0.17 mmol) in DMF (0.5 mL) was added and the mixture was stirred for 1.5 hours. H_2O/ ACN 1:1 (30 mL, v/v) was added to the reaction mixture and it was freeze-dried overnight. The resulting solid was treated

with a TFA/H_2O/TIPS 9.5:2.5:2.5 (5 mL, v/v/v) mixture for 2 hours. Volatile acids and solvents were removed under a stream of nitrogen. The crude peptide was then precipitated by adding diethyl ether (45 mL), the solid was recovered by centrifugation, re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried. The peptide was purified by semi-preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) and freeze-dried to yield **29** (16 mg, 17.3 μmol, 32%) as a white solid.

7.3.2. Synthesis of peptide methyl esters

Following the SPPS methodology outlined in section 7.2.1, the crude peptide as the C-terminal carboxylic acid was dissolved in a solution of 1:1 trimethyl orthoformate/methanol and conc. HCl (5% v/v) and stirred for 48 hours. The solvents were then removed at reduced pressure. The product was re-dissolved in H_2O/ACN 1:1 (v/v) and freeze-dried overnight. The peptide was purified using preparative HPLC (MeOH/water) and freeze-dried to yield the desired peptide methyl ester as a white solid.

Compound	Isolated yield[ix]
34	24.1 mg, 32.6 μmol, 33%
35	39.2 mg, 53.0 μmol, 53%
36	22.4 mg, 30.0 μmol 30%
38	14.9 mg, 20.0 μmol, 20%
39	3 mg, 1.8 μmol, 10%
40	0.5 mg, 0.3 μmol, 2%
41	0.3 mg, 0.2 μmol, 1%

7.3.3. Synthesis of peptide *N*-hydroxysuccinimide esters

49[x]

Following the SPPS methodology outlined in section 7.2.1, the crude peptide as the C-terminal carboxylic acid (30 mg, 72 μmol) was dissolved in DMF (1 mL) and cooled to 0 °C. A solution of EDC·HCl (166 mg, 864 μmol, 12 equivalents) and NHS (124 mg, 1080 μmol, 15 equivalents) in DMF (2 mL) was then added and the resulting mixture stirred for 24 hours at 0 °C.[107] The reaction was

[ix] Synthesis of **34** and **36** was carried out by Kelly Chu.
[x] Synthesis of peptide **49** has been previously reported by Danny Burke.

quenched with cold water (20 mL, 0.1% v/v TFA) and lyophilised immediately. The crude peptide was purified by preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) to afford **49** as a white solid (14.8 mg, 29 μmol, 40%).

52 and **53**

Following the SPPS methodology outlined in section 7.2.1, the crude peptide as the C-terminal carboxylic acid (20 mg, 28 μmol) was dissolved in DMF (1 mL) and cooled to 0 °C. A solution of EDC·HCl (64 mg, 0.37 mmol) and NHS (48 mg, 0.42 mmol) in DMF (2 mL) was then added and the resulting mixture stirred for 24 hours at 0 °C.[107] The reaction was quenched with cold water (20 mL, 0.1% v/v TFA) and lyophilised immediately. The crude peptide was purified by preparative HPLC (acetonitrile (0.1% v/v TFA)/water (0.1% v/v TFA)) to afford the desired peptide *N*-hydroxysuccinimide esters as white solids.

Compound	Isolated yield
52	1.4 mg, 1.7 μmol 6%
53	0.9 mg, 1.1 μmol, 4%

7.3.4. Synthesis of peptide $(EG)_3$ ester 43

Following the SPPS methodology outlined in section 7.2.1, the crude peptide as the C-terminal carboxylic acid (10 mg, 5.4 μmol) was dissolved in triethylene glycol monomethyl ether (10 mL) containing conc. HCl (5% v/v) and stirred for 48 hours at 60 °C. The reaction was cooled to room temperature and water (20 mL) was added to precipitate the peptide, then it was filtered, resuspended in H_2O/ACN 1:1 (v/v) and freeze-dried overnight. The crude peptide was purified by preparative HPLC (MeOH/water) to afford **43** as a white solid (0.8 mg, 0.4 μmol, 7%).

7.4. Native chemical ligation procedures

7.4.1. Template-free native chemical ligation experiments

Experiments were initiated by preparing H_2O/ACN solutions containing equimolar amounts (0.5-2 mM) of **19** and the appropriate electrophilic peptide fragment (**22-25**) with tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl, 5-20 mM) as reducing agent under N_2. After equilibrating for ~10 minutes, reactions were initiated by the addition of MOPS buffer (0.5 mL) at an appropriate pH, yielding a total volume of 1 mL (including up to 40% in volume of ACN). The experiments were carried out in Young's

tubes. A volume of 0.2 mL was removed at the studied time points and immediately analysed by analytical HPLC, using a Phenomenex C4 column at a 1.0 mL/min flow rate, with a gradient of 0 to 60% Acetonitrile in 0.1 M NH_4HCO_3 (aq, pH = 8) over 20 min, monitoring at 220 nm and 270 nm channels. If not injected immediately, the sample was stored frozen until analysis.

7.4.2. Ashkenasy's system template-seeded native chemical ligation experiments

Water/ACN (1:1) solutions of **1** and **28** (1 and 2 mM, respectively) were mixed in the presence of TCEP·HCl and sonicated for 10 minutes. The mixture was then diluted with MOPS buffer and sonicated for another 10 minutes. After equilibrating for 20 minutes, reactions were initiated by adding a water/ACN (1:1) solution of the electrophilic fragment **30** (2 mM). The final conditions were: **28** and **30** at 250 µM, **1** at 100 µM, TCEP 5 mM, 17.5% volume of ACN and pH = 6.8-7.1. The experiments were carried out in a 5 mL round bottom flask, where a 150 µL sample was taken at the studied time points, immediately injected into the Gilson analytical HPLC, equipped with an automatic injector, using a Phenomenex C4 column at a 1.0 mL/min flow rate, with a gradient of 0 to 50% ACN in 0.1 M NH_4HCO_3 (aq, pH = 8) over 40 min, monitoring at 220 nm and 270 nm channels. If not injected immediately, the samples were stored frozen until analysis. The concentration of each compound was calculated by comparison with the HPLC peak of an ABA-labelled peptide (**Std**) with known concentration. All experiments were repeated at least three times, and error bars for repeating experiments were found to be in the order of ± 7 %.

7.4.3. Template recovery experiments

A 1 mM water/ACN (1:1) solution of **1** in the presence of TCEP·HCl was diluted with MOPS buffer. The final conditions were: **1** at 100 µM, TCEP 5 mM, 17.5% volume of ACN and pH = 7.1.The mixture was sonicated for 10 minutes. The experiments were carried out in a 5 mL round bottom flask, where a 150 µL sample was taken at the studied time points. Samples were either stored frozen or immediately injected into a Gilson analytical HPLC, using a Phenomenex C4 column at a 1.0 mL/min flow rate, with a gradient of 0 to 50% ACN in 0.1 M NH_4HCO_3 (aq, pH = 8) over 40 min, monitoring at 220 nm and 270 nm channels. The concentration of **1** was calculated by comparison with the HPLC peak of an ABA-labelled peptide (**Std**) with known concentration. All experiments were repeated at least three times, and error bars for repeating experiments (including frozen samples) were found to be in the order of ± 3%.

7.5. *N*-hydroxysuccinimide coupling procedures

7.5.1. Competition coupling experiments in the linear system

Nucleophilic fragments **47** and **48** were dissolved in methanol and sonicated for 5 minutes. 2 equivalents of DIPEA were added and the mixture was stirred for 5 minutes. Electrophilic ester **49** was dissolved in methanol and sonicated for 5 minutes. Both solutions were then mixed and stirred for 5 minutes to initiate the reaction. The final peptide concentration was 1 mM, the reactions were carried out in a 200 µL scale. 80 µL samples were taken after 1 and 4 hours of initiating the reactions and quenched with 120 µL of water (0.1% TFA) to yield a total volume of 200 µL, the samples were immediately analysed by analytical HPLC in a Waters equipment, using a XSELECT CSH C18 5µm (4.6 × 150 mm) column at a 1.0 mL/min pump rate, with a gradient of 10 to 30% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 5 minutes, then an isocratic elution of 30% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 35 minutes and a gradient of 30 to 100% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 5 minutes, monitoring at 220 nm. The ratio of the areas under the peaks corresponding to **45** and **46** was then calculated.

7.5.2. Template seeded couplings

Nucleophilic fragment **47** and template **45** were dissolved in the appropriate solvent (or mixture of solvents) and sonicated for 5 minutes. 2 equivalents of DIPEA were added and the mixture was stirred for 5 minutes. Electrophilic fragment **49** was dissolved in the appropriate solvent (or mixture of solvents) and sonicated for 5 minutes. Both solutions were then mixed and stirred for 5 minutes to initiate the reaction. The reactions were carried out in a 400 µL scale. Samples were taken after 0 min, 30 min, 1.5 h, 3 h, and 6 h of initiating the reaction. At each time point 80 µL samples were taken and mixed with 80 µL of a 250 µM solution of **Std** in methanol to yield a total volume of 200 µL, where the concentration of **30** was kept constant. The samples were analysed by analytical HPLC in a Waters equipment, using a XSELECT CSH C18 5µm (4.6 × 150 mm) column at a 1.0 mL/min pump rate, with a gradient of 10 to 30% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 5 minutes, then an isocratic elution of 30% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 35 minutes and a gradient of 30 to 100% acetonitrile (0.1% v/v TFA) in water (0.1% v/v TFA) over 5 minutes, monitoring at 220 nm. The concentration of **45** was calculated by using a calibration plot built from the correlation of the ratio of areas corresponding to **45** and an internal standard **Std** with the ratio of their concentration (Figure 7.1).

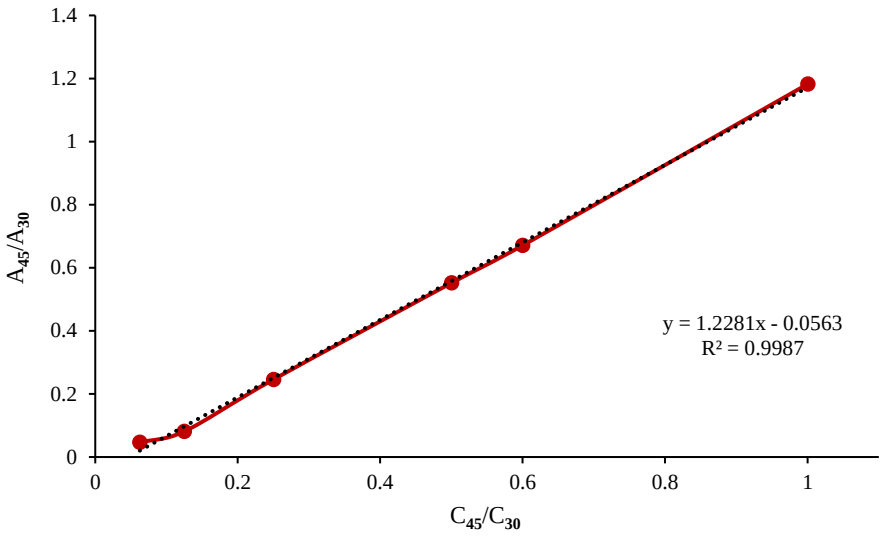


Figure 7.1: Calibration plot of the ratio of response vs ratio of concentration of **45**/**Std**. The equation and R^2 value of the fitted line plot are embedded.

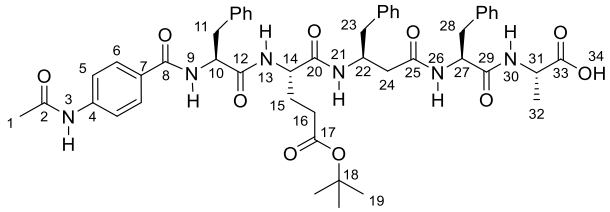
7.5.3. Competition coupling experiments in the β-hairpin system

Nucleophilic fragment **51** was dissolved in the appropriate solvent and sonicated for 5 minutes. 1.2 equivalents of DIPEA were added and the mixture was stirred for 5 minutes. Electrophilic esters **52** and **53** were dissolved in the appropriate solvent and sonicated for 5 minutes. Both solutions were then mixed and stirred for 5 minutes to initiate the reaction. The final peptide concentration was 250 µM. The reactions were carried out in a 200 µL scale. After 4 hours of initiating the reaction, 80 µL samples were taken and quenched with 80 µL of MeOH/water (0.1% TFA) 1:1 to yield a total volume of 160 µL. The samples were immediately analysed by HPLC in a Waters equipment, using a XSELECT CSH C18 5µm (4.6 × 150 mm) column at a 1.0 mL/min pump rate, with a gradient of 70 to 100% methanol in water over 40 minutes, monitoring at 220 nm. The ratio of the areas under the peaks corresponding to **54** and **55** was then calculated.

7.6. NMR characterisation data for selected peptide sequences

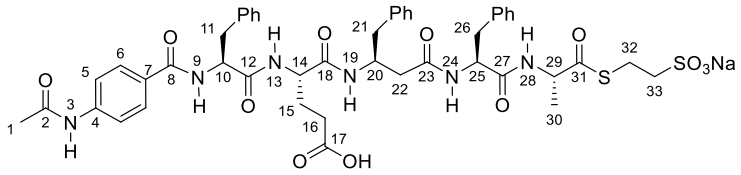
7.6.1. Chapter 2

21a



^{1}H NMR (500 MHz, DMSO-d_6): δ 1.29 (3H, d, *J* = 7.3 Hz, C(32)-$\underline{H}_3$), 1.36 (9H, s, 3 × (C(19)-$\underline{H}_3$)), 1.60-1.83 (2H, m C(15)-$\underline{H}_2$), 2.04-2.26 (7H, m, C(1)-$\underline{H}_3$,C(16)-$\underline{H}_2$, C(24)-$\underline{H}_2$), 2.37-2.46 (2H, m, C(23)-$\underline{H}_2$), 2.72 (1H, dd, *J* = 10.4, 13.9 Hz, C(28)-$\underline{H}$), 2.96 (1H, dd, *J* = 10.7, 13.9 Hz, C(11)-$\underline{H}$), 3.01-3.10 (2H, m, C(11)-$\underline{H}$', C(28)-$\underline{H}$'), 4.06-4.22 (3H, m, C(14)-$\underline{H}$, C(22)-$\underline{H}$, C(31)-$\underline{H}$), 4.58-4.70 (2H, m, C(10)-$\underline{H}$, C(27)-$\underline{H}$), 6.97-7.37 (15H, m, Phe residues Aryl C-$\underline{H}$), 7.61 (2H, d, *J* = 6.4 Hz, C(5)-$\underline{H}$), 7.70-7.77 (3H, m, C(6)-$\underline{H}$, N(21)-$\underline{H}$), 8.09-8.21 (2H, m, N(13)-$\underline{H}$, N(26)-$\underline{H}$), 8.29 (1H, s, N(30)-$\underline{H}$), 8.44 (1H, d, *J* = 8.2 Hz, N(9)-$\underline{H}$), 10.14 (1H, s, N(3)-$\underline{H}$), 12.57 (1H, br s, O(34)-$\underline{H}$). **^{13}C NMR** (125 MHz, DMSO-d_6): δ 17.2 ($\underline{C}$(32)-H_3), 24.1 ($\underline{C}$(1)-H_3), 27.4 ($\underline{C}$(15)-H_2), 27.7 (($\underline{C}$(19)-H_3)$_3$), 31.1 ($\underline{C}$(16)-H_2), 36.9 ($\underline{C}$(11)-H_2), 37.5 ($\underline{C}$(28)-H_2), 38.4 ($\underline{C}$(23)-H_2), 39.8 ($\underline{C}$(24)-H_2, assigned by HSQC), 47.5 ($\underline{C}$(22)-H), 47.7 ($\underline{C}$(31)-H), 52.0 ($\underline{C}$(14)-H), 53.4 ($\underline{C}$(27)-H), 54.9 ($\underline{C}$(10)-H), 79.6 ($\underline{C}$18), 117.9 ($\underline{C}$(5)-H), 125.8, 126.1, 126.2, 127.9, 128.0, 129.2 (Phe residues Aryl $\underline{C}$-H), 128.2 ($\underline{C}$(7)), 128.3 ($\underline{C}$(6)-H), 137.9, 138.4, 138.5 (Phe residues Aryl $\underline{C}$), 142.0 ($\underline{C}$(4)), 165.9 ($\underline{C}$(8)=O), 168.7 ($\underline{C}$(2)=O), 169.6 ($\underline{C}$(25)=O), 169.9 ($\underline{C}$(20)=O), 170.9, 171.4 ($\underline{C}$(12/29)=O), 171.7 ($\underline{C}$(17)=O), 173.9 ($\underline{C}$(33)=O).

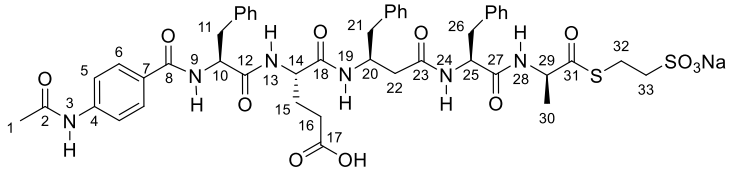
23



^{1}H NMR (500 MHz, DMSO-d_6): δ 1.29 (3H, d, *J* = 7.2 Hz, C(30)-$\underline{H}_3$), 1.65-1.85 (2H, m, C(15)-$\underline{H}_2$), 2.05 (3H, s, C(1)-$\underline{H}_3$), 2.07-2.23 (4H, m, C(16)-$\underline{H}_2$, C(22)-$\underline{H}_2$), 2.37-2.43 (2H, m, C(21)-$\underline{H}_2$), 2.54 (2H, t, *J* = 8.1 Hz, C(33)-$\underline{H}_2$), 2.71-2.79 (1H, m, C(26)-$\underline{H}$), 2.91-3.02 (3H, m, C(11)-$\underline{H}$, C(32)-$\underline{H}_2$), 3.05 (1H, dd, *J* = 3.9, 13.8 Hz, C(11)-$\underline{H}$'), 3.17 (1H, dd, *J* = 4.1, 14 Hz, C(26)-$\underline{H}$'), 4.07-4.18 (2H, m, C(14)-$\underline{H}$, C(20)-$\underline{H}$), 4.40 (1H, p, *J* = 7.2 Hz, C(29)-$\underline{H}$), 4.62-4.73 (2H, m, C(10)-$\underline{H}$, C(25)-$\underline{H}$), 6.97-7.36 (15H, m, Phe residues Aryl C-$\underline{H}$), 7.61 (2H, d, *J* = 9 Hz, C(5)-$\underline{H}$), 7.72 (2H, d, *J* = 9 Hz, C(6)-$\underline{H}$), 7.75 (1H, d, *J* = 8.1 Hz, N(19)-$\underline{H}$), 8.08 (1H, d, *J* = 8.0 Hz, N(13)-$\underline{H}$), 8.25 (1H, d, *J* = 8.7 Hz, N(24)-$\underline{H}$), 8.41 (1H, d,

J = 8.3 Hz, N(9)-H), 8.77 (1H, d, *J* = 7.2 Hz, N(28)-H), 10.13 (1H, s, N(3)-H). **^{13}C NMR** (125 MHz, DMSO-d_6): δ 17.3 (C(30)-H_3), 24.10, 24.14 ((C(1)-H_3/C(32)-H_2), 27.5 ((C(15)-H_2), 30.0 (C(16)-H_2), 37.0 (C(11)-H_2), 37.3 (C(26)-H_2), 38.4 (C(21)-H_2), 39.7 (C(22)-H_2, assigned by HSQC), 47.4 (C(20)-H), 50.8 (C(33)-H_2), 52.1 (C(14)-H), 53.3 (C(25)-H), 54.81 (C(29)-H), 54.89 (C(10)-H), 117.9 (C(5)-H), 125.8, 126.2, 127.89, 127.93, 128.0, 128.25, 129.1 (Phe residues Aryl C-H), 128.22 (C(4)), 129.2 (C(6)-H), 137.9, 138.4, 138.5 (Phe residues Aryl C), 142.0 (C(7)), 165.8 (C(8)=O), 168.7 (C(2)=O), 169.6 (C(23)=O), 169.9 (C(18)=O), 171.3, 171.4 (C(12/27)=O), 173.9 (C(17)=O), 201.7 (C(31)=O).

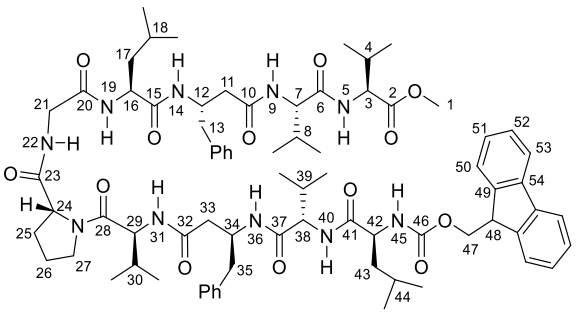
23'



^{1}H NMR (500 MHz, DMSO-d_6): δ 1.18 (3H, d, *J* = 7.3 Hz, C(30)-H_3), 1.67-1.87 (2H, m, C(15)-H_2), 2.05 (3H, s, C(1)-H_3), 2.07-2.30 (4H, m, C(16)-H_2, C(22)-H_2), 2.46-2.58 (4H, m, C(21)-H_2, C(33)-H_2, assigned by COSY), 2.78 (1H, dd, *J* = 9.4, 13.8 Hz, C(26)-H), 2.93-3.09 (5H, m, C(26)-H', C(11)-H_2, C(32)-H_2), 4.12-4.20 (2H, m, C(14)-H, C(20)-H), 4.35 (1H, p, *J* = 7.2 Hz, C(29)-H), 4.62-4.69 (2H, m, C(10)-H, C(25)-H), 6.96-7.38 (15H, m, Phe residues Aryl C-H), 7.61 (2H, d, *J* = 8.7 Hz, C(5)-H), 7.73 (2H, d, *J* = 8.7 Hz, C(6)-H), 7.82 (1H, d, *J* = 8.2 Hz, N(19)-H), 8.10 (1H, d, *J* = 7.9 Hz, N(13)-H), 8.23 (1H, d, *J* = 8.6 Hz, N(24)-H), 8.42 (1H, d, *J* = 8.3 Hz, N(9)-H), 8.54 (1H, d, *J* = 7.3 Hz, N(28)-H), 10.14 (1H, s, N(3)-H). **^{13}C NMR** (125 MHz, DMSO-d_6): δ 17.2 (C(30)-H_3), 24.1, 24.2 ((C(1)-H_3/C(32)-H_2), 27.6 ((C(15)-H_2), 30.0 (C(16)-H_2), 37.0 (C(11)-H_2), 37.7 (C(26)-H_2), 38.5 (C(21)-H_2), 40.0 (C(22)-H_2, assigned by HSQC), 47.4 (C(20)-H), 50.8 (C(33)-H_2), 52.1 (C(14)-H), 53.5 (C(25)-H), 54.7 (C(29)-H), 55.0 (C(10)-H), 117.9 (C(5)-H), 125.9, 126.16, 126.19, 127.94, 127.98, 128.0, 129.1, 129.18, 129.19 (Phe residues Aryl C-H), 128.2 (C(4)), 128.3 (C(6)-H), 137.7, 138.4, 138.5 (Phe residues Aryl C), 142.0 (C(7)), 165.8 (C(8)=O), 168.7 (C(2)=O), 169.6 (C(23)=O), 170.0 (C(18)=O), 171.2, 171.4 (C(12/27)=O), 173.9 (C(17)=O), 201.3 (C(31)=O).

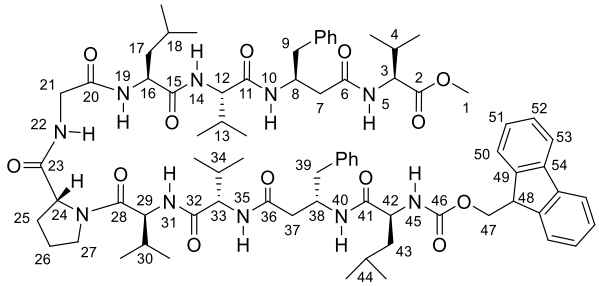
7.6.2. Chapter 3

34



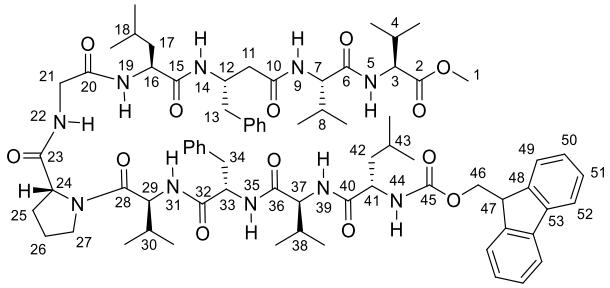
^{1}H NMR (500 MHz, chloroform-d): δ 0.77-1.09 (36H, m, Val and Leu residues 12 × C-H$_3$), 1.45-1.53 (1H, m, C(17)-H), 1.50-1.85 (6H, m, C(17)-H$_2$, C(18)-H, C(43)-H$_2$, C(44)-H), 1.92-2.07 (3H, m, C(8)-H, C(25)-H, C(39)-H), 2.08-2.28 (7H, m, C(4)-H, C(11)-H$_2$, C(25)-H', C(26)-H$_2$, C(30)-H), 2.31-2.40 (1H, m C(33)-H), 2.55-2.70 (2H, m, C(13)-H, C(35)-H), 2.78 (1H, dd, *J* = 14.2, 4.1 Hz, C(13)-H'), 2.86-3.05 (2H, m, C(33)-H', C(35)-H'), 3.6-3.82 (6H, m, C(1)-H$_3$, C(21)-H, C(27)-H,), 4.0-4.13 (1H, m, C(21)-H'), 4.17-4.46 (7H, m, C(12)-H, C(16)-H, C(24)-H, C(42)-H, C(47)-H$_2$, C(48)-H), 4.62 (2H, apparent br s, C(3)-H, C(38)-H), 4.84 (2H, apparent br s, C(29)-H, C(34)-H), 5.06 (1H, br s, C(7)-H), 6.28-6.94 (2H, m, N(22)-H, N(45)-H), 6.90-7.22 (11H, m, N(9)-H, N(14)-H, N(40)-H, β^{3}hPhe residues Aryl C-H), 7.22-7.32 (3H, m, C(51)-H, β^{3}hPhe residues Aryl C-H), 7.34-7.42 (2H, m, C(52)-H), 7.61 (2H, dd, *J* = 7.5, 4.4 Hz, C(50)-H), 7.75 (2H, d, *J* = 7.4 Hz, C(53)-H), 7.85-8.08 (2H, m, N(19)-H, N(36)-H), 8.26 (1H, br s, N(31)-H), 8.53 (1H, br s, N(5)-H). **^{13}C NMR** (125 MHz, chloroform-d): δ 18.9, 19.0, 19.3, 22.0, 22.8, 23.3 (Val and Leu residues 12 × C-H$_3$), 24.9 (C(18)-H/C(44)-H, assigned by HSQC), 25.3 (C(25)-H$_2$), 29.1 (C(26)-H$_2$), 31.5, 31.6, 32.1, 32.6 (C(4)-H/C(8)-H/C(30)-H/C(39)-H), 38.5 (C(13)-H$_2$), 38.9 ((C(35)-H$_2$), 39.3 (C(17)-H$_2$), 40.3 (C(33)-H$_2$), 40.5 (C(11)-H$_2$), 41.6 (C(43)-H$_2$), 43.3 (C(21)-H$_2$), 47.1, 47.0 (C(29)-H/C(34)-H), 47.4 (C(24)-H), 47.8 (C(42)-H), 48.2 (C(27)-H$_2$), 51.5 (C(16)-H), 52.3 (C(1)-H$_3$), 54.1 (C(12)-H), 57.8, 57.6 (C(3)-H/C(7)-H/C(38)-H, assigned by HSQC), 61.3 (C(48)-H), 67.1 (C(47)-H$_2$), 120.1 (C(53)-H), 125.3, 125.4 (C(50)-H/C(50)-H'), 126.1, 126.9, 127.1, 127.8, 128.0, 128.8, 129.5, 130.4 (C(51)-H, (C(52)-H), β^{3}hPhe residues Aryl C-H), 136.7, 137.5 (β^{3}hPhe residues Aryl C), 141.4 (C(54)), 144.0, 144.2 (C(49)/C(49)'), 156.6 (C(46)=O), 169.1 (C(20)=O), 169.9, 170.0, 170.5, 171.1, 171.8 (C(6)=O/C(10)=O/C(15)=O/C(37)=O/C(41)=O), 172.1 (C(23)=O), 173.1 (C(2)=O), 174.7 (C(28)=O, assigned by HMBC), 175.5 (C(32)=O, assigned by HMBC).

35



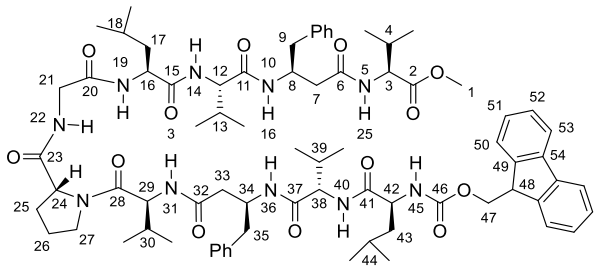
[1]H NMR (500 MHz, chloroform-d): δ 0.77-1.06 (36H, m, Val and Leu residues 12 × C-H$_3$), 1.42-1.76 (5H, m, C(17)-H, C(18)-H, C(43)-H$_2$, C(44)-H), 1.86 (1H, m, C(17)-H'), 1.92-2.07 (3H, m, C(12)-H, C(25)-H, C(34)-H), 2.07-2.39 (8H, m, C(4)-H, C(7)-H, C(25)-H', C(26)-H$_2$, C(30)-H, C(37)-H$_2$), 2.61-2.79 (3H, m, C(7)-H', C(9)-H, C(39)-H), 2.80-2.90 (2H, m, C(9)-H', C(39)-H'), 3.57-3.74 (4H, m, C(1)-H$_3$, C(27)-H), 3.76-3.91 (2H, m, C(21)-H, C(27)-H'), 4.08-4.28 (4H, m, C(21)-H', C(42)-H, C(47)-H, C(48)-H), 4.30-4.52 (5H, m, C(12)-H, C(16)-H, C(24)-H, C(38)-H, C(47)-H'), 4.60 (1H, br s, C(29)-H), 4.64-4.80 (2H, m, C(8)-H, C(3)-H), 4.91 (1H, br s, C(33)-H), 5.96 (1H, br s, N(59)-H), 6.31 (1H, br s, N(22)-H), 6.82 (3H, apparent br s, N(14)-H, N(35)-H, N(40)-H), 7.00-7.17 (10H, m, β^{3}hPhe residues Aryl C-H), 7.26-7.33 (2H, m, C(51)-H), 7.39 (2H, t, *J* = 7.5 Hz, C(52)-H), 7.54-7.61 (2H, m, C(50)-H), 7.62-7.79 (4H, m, N(5)-H, N(19)-H, C(53)-H), 7.91 (1H, br s, N(10)-H), 8.67 (1H, br s, N(31)-H). **[13]C NMR** (125 MHz, chloroform-d): δ 18.5, 18.7, 18.8, 18.96, 18.98, 19.01, 19.2, 22.0 22.2, 22.7, 22.8, 23.1 (Val and Leu residues C-H$_3$), 24.7 (C(44)-H), 24.8 (C(18)-H), 25.1 (C(25)-H), 29.0 (C(26)-H$_2$), 30.9 (C(13)-H), 31.1 (C(30)-H), 31.7 (C(4)-H), 32.0 (C(34)-H), 39.1, 39.4 (C(9)-H$_2$/C(39)-H$_2$), 39.8 (C(7)-H$_2$), 40.3 (C(17)-H$_2$), 40.5 (C(37)-H$_2$), 41.0 ((C(43)-H$_2$), 43.1 (C(21)-H), 47.2 (C(48)-H), 47.4 (C(8)-H), 48.0 (C(27)-H$_2$), 48.5 (C(24)-H), 52.0 (C(1)-H$_3$), 52.1 (C(16)-H), 53.6 (C(42)-H), 56.5 (C(29)-H), 57.2 (C(3)-H), 58.8 (C(33)-H), 61.0 (C(38)-H), 67.0 (C(47)-H$_2$), 120.0 (C(53)-H), 125.0, 125.2 (C(50)-H/C(50)-H'), 126.1, 126.6 (β^{3}hPhe residues Aryl C-H), 127.0, 127.1 (C(51)-H/C(51)-H'), 127.8 (C(52)-H), 128.0, 128.4, 129.2, 130.0 (β^{3}hPhe residues Aryl C-H), 137.16, 137.68 (β^{3}hPhe residues Aryl C), 141.3 (C(54)), 143.79, 143.84 (C(49)/C(49)'), 156.3 (C(46)=O), 170.58, 170.63 (C(6)=O)/C(11)=O), 171.6 (C(23)=O), 173.0 (C(2)=O), 168.90, 168.91, 170.1, 171.4, 171.8, 172.3 (C(15)=O/C(20)=O/C(28)=O/C(32)=O/ C(36)=O/C(41)=O,).

36



^{1}H NMR (500 MHz, methanol-d_3): δ 0.75-0.98 (m, 36H, Val and Leu residues 12 × C-H$_3$), 1.22-1.44 (3H, m, C(17)-H C(42)-H$_2$, assigned by HSQC), 1.57 (1H, m, C(43)-H), 1.70 (1H, m, C(18)-H), 1.85 (1H, ddd, *J* = 13.6, 11.4, 4.4 Hz, C(17)-H'), 1.94-2.24 (9H, m, C(4)-H, C(8)-H, C(11)-H, C(25)-H$_2$, C(26)-H$_2$, C(30)-H, C(38)-H), 2.49 (1H, dd, *J* = 14.8, 4.9 Hz, C(11)-H'), 2.64 (1H, dd, *J* = 13.6, 7.7 Hz, C(13)-H), 2.80 (1H, dd, *J* = 13.2, 6.2 Hz, C(13)-H'), 2.89 (1H, dd, *J* = 14.0, 9.3 Hz, C(34)-H), 3.14 (1H, dd, *J* = 14.0, 4.9 Hz, C(34)-H'), 3.59-3.71 (5H, m, C(1)-H$_3$, C(27)-H$_2$), 3.76 (1H, dd, *J* = 17.1, 3.9 Hz, C(21)-H), 3.97 (1H, dd, *J* = 17.1, 4.0 Hz, C(21)-H'), 4.16-4.48 (10H, m, C(3)-H, C(7)-H, C(16)-H, C(24)-H, C(29)-H, C(37)-H, C(41)-H, C(46)-H$_2$, C(47)-H), 4.56 (1H, t, *J* = 7.0 Hz, C(12)-H), 5.18 (1H, td, *J* = 8.8, 4.7 Hz, C(33)-H), 6.92-7.24 (11H, m, N(44)-H (assigned by TOCSY), β^{3}hPhe and Phe residues Aryl C-H), 7.25-7.31 (2H, m, C(50)-H), 7.37 (2H, td, *J* = 7.5, 1.0 Hz, C(51)-H), 7.61 (2H, t, *J* = 6.9 Hz, C(49)-H), 7.78 (2H, d, *J* = 7.5 Hz, C(52)-H), 8.05 (2H, apparent d, *J* = 8.3 Hz, N(14)-H, N(19)-H), 8.10 (1H, d, *J* = 8.8 Hz, N(9)-H), 8.18 (2H, apparent d, *J* = 9.2 Hz, N(22)-H, N(39)-H, assigned by TOCSY), 8.31 (1H, d, *J* = 7.6 Hz, N(5)-H), 8.36 (1H, d, *J* = 8.6 Hz, N(35)-H), 8.54 (1H, d, *J* = 8.5 Hz, N(31)-H). **^{13}C NMR** (125 MHz, methanol-d_3): δ 18.4, 18.9, 19.1, 19.2, 19.4, 19.7, 19.7, 19.8, 21.4, 21.9, 23.6, 23.6 (Val and Leu residues C-H$_3$), 25.5 (C(18)-H), 25.7 (C(43)-H), 25.9 (C(26)-H$_2$), 30.0 (C(25)-H$_2$), 31.3 (C(30)-H), 32.1 (C(4)-H), 32.4 (C(8)-H), 32.9 (C(38)-H), 39.0 (C(34)-H$_2$), 40.7 (C(11)-H$_2$), 41.4 (C(13)-H$_2$), 42.1 (C(17)-H$_2$), 42.5 (C(43)-H$_2$), 43.8 (C(21)-H$_2$), 48.4 (C(47)-H, assigned by HSQC), 48.9 (C(27)-H$_2$, assigned by HSQC), 49.6 (C(12)-H, assigned by HSQC), 52.1 (C(1)-H$_3$), 53.0 (C(16)-H), 54.9 (C(37)-H), 55.5 (C(33)-H), 58.3 (C(29)-H), 59.3, 59.58 (C(7)-H/C(37)-H), 59.62 (C(3)-H), 62.2 (C(24)-H), 67.8 (C(46)-H$_2$), 120.8 (C(52)-H), 126.1, 126.1 (C(49)-H/ C(49)-H'), 127.3, 127.4 (β^{3}hPhe and Phe residues Aryl C-H), 128.02, 127.98 (C(51)-H/C(51)-H'), 128.6 (C(50)-H), 129.1, 129.3, 130.3, 130.4 (β^{3}hPhe and Phe residues Aryl C-H), 138.7, 139.0 (β^{3}hPhe and Phe residues Aryl C), 142.4, 142.5 (C(53)/C(53)'), 144.9, 145.4 (C(48)/C(48)'), 158.0 (C(45)=O), 171.5 (C(20)=O), 172.2 (C(28)=O), 172.9 (C(36)=O), 173.0 (C(10)=O), 173.4 (C(2)=O), 174.0 (C(15)=O), 174.1 (C(6)=O), 174.2 (C(32)=O), 174.8 (C(23)=O), 175.1 (C(40)=O).

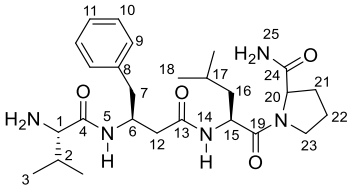
38



[1]H NMR (500 MHz, chloroform-d): δ 0.77-0.99 (36H, m, Val and Leu residues 12 × C-H$_3$), 1.45-1.53 (1H, m, C(17)-H), 1.54-1.76 (5H, m, C(17)-H', C(18)-H, C(43)-H$_2$, C(44)-H), 1.94-2.17 (8H, m, C(4)-H, C(13)-H, C(25)-H$_2$, C(26)-H$_2$, C(30)-H, C(39)-H), 2.24-2.35 (2H, m, C(7)-H, C(33)-H), 2.46 (1H, dd, *J* = 15.1, 5.0 Hz, C(7)-H'), 2.54 (1H, d, *J* = 13.7 Hz, C(33)-H'), 2.77-2.85 (3H, m, C(9)-H, C(35)-H$_2$), 2.97 (1H, dd, *J* = 13.5, 6.0 Hz, C(9)-H'), 3.56-3.64 (1H, m, C(27)-H), 3.70 (3H, s, C(1)-H$_3$), 3.82 (1H, dd, *J* = 17.1, 5.7 Hz, C(21)-H), 3.89-4.03 (2H, m, C(21)-H', C(27)-H'), 4.13 (1H, br s, C(38)-H), 4.19-4.28 (3H, m, C(12)-H, C(42)-H, C(48)-H), 4.29-4.36 (2H, m, C(24)-H, C(47)-H), 4.37-4.48 (4H, m, C(8)-H, C(16)-H, C(29)-H, C(47)-H'), 4.50-4.60 (2H, m, C(3)-H, C(34)-H), 5.97 (1H, br s, N(45)-H), 6.26 (1H, d, *J* = 8.0 Hz, N(5)-H), 6.74 (1H, br s, N(40)-H), 6.83 (1H, br s, N(22)-H), 7.03 (1H, br s, N(14)-H), 7.11-7.32 (11H, m, N(36)-H, β^{3}hPhe residues Aryl C-H), 7.33-7.42 (3H, m, C(52)-H, N(31)-H), 7.51 (1H, br s, N(19)-H), 7.59 (3H, apparent d, *J* = 7.5 Hz, N(10)-H, C(50)-H), 7.75 (2H, d, *J* = 7.6 Hz, C(53)-H). **[13]C NMR** (125 MHz, chloroform-d): δ 18.09, 18.13, 18.3, 18.7, 19.0, 19.1, 19.3, 19.4, 22.1, 22.8, 23.2, 23.3 (Val and Leu residues 12 × C-H$_3$), 24.6, 24.8 (C(18)-H/ C(44)-H), 25.0 (C(25)-H$_2$), 29.3 (C(26)-H$_2$), 30.8, 30.9, 31.0 (C(4)-H/C(30)-H/C(39)-H), 31.2 (C(13)-H), 38.4 (C(7)-H$_2$), 39.5 ((C(33)-H$_2$), 39.9, 40.2 (C(9)-H$_2$/C(39)-H), 40.5 (C(17)-H$_2$), 41.4 (C(43)-H$_2$), 47.3 (C(48)-H), 47.98, 48.03 (C(27)-H$_2$/C(34)-H), 48.7 (C(8)-H), 52.3 (C(1)-H$_3$, C(16)-H, assigned by HSQC), 53.9 (C(42)-H), 57.0 (C(29)-H), 57.2 (C(3)-H), 59.1 (C(12)-H), 59.5 (C(38)-H), 61.3 (C(24)-H), 67.2 (C(47)-H$_2$), 120.1 (C(53)-H), 125.3 (C(50)-H), 126.7, 126.8 (β^{3}hPhe residues Aryl C-H), 127.2 (C(51)-H), 127.9 (C(52)-H), 128.5, 128.7, 129.4, 129.6 (β^{3}hPhe residues Aryl C-H), 137.8, 138.1 (β^{3}hPhe residues Aryl C), 141.4 (C(54)), 143.9, 144.0 (C(49)/C(49)'), 156.6 (C(46)=O), 169.8, 172.0 (C(20)=O/C(23)=O), 170.7 (C(15)=O), 171.03 (C(32)=O), 171.07 (C(28)=O), 171.1 (C(6)=O), 171.0, 172.1, 172.5 (C(11)=O/C(37)=O/C(41)=O)/172.6 (C(2)=O).

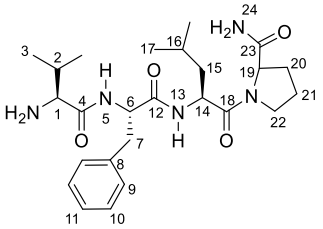
7.6.3. Chapter 4[xi]

47



^{1}H NMR (500 MHz, methanol-d_3): δ 0.88-0.98 (12H, m, 2 × C(3)-H$_3$, 2 × C(18)-H$_3$), 1.53-1.60 (2H, m, C(16)-H$_2$), 1.65-1.75 (1H, m, C(17)-H), 1.91-2.22 (4H, m, C(2)-H, C(21)-H, C(22)-H$_2$), 2.16-2.24 (1H, m, C(21)-H'), 2.43 (1H, dd, *J* = 5.1, 14.8 Hz, C(12)-H), 2.56 (1H, dd, *J* = 9.3, 14.8 Hz, C(12)-H'), 2.85 (1H, dd, *J* = 6.9, 13.7 Hz, C(7)-H), 2.98 (1H, dd, *J* = 7.1, 13.7 Hz, C(7)-H'), 3.49 (1H, d, *J* = 5.5 Hz, C(1)-H), 3.59-3.66 (1H, m, C(23)-H), 3.80-3.86 (1H, m, C(23)-H'), 4.33-4.41 (2H, m, C(6)-H, C(20)-H), 4.61 (1H, q, *J* = 7.4 Hz, C(15)-H), 6.96 (1H, s, N(25)-H), 7.16-7.31 (5H, m, 2 × C(9)-H, 2 × C(10)-H, C(11)-H), 7.53 (1H, s, N(25)-H'), 8.11 (1H, d, *J* = 7.4 Hz, N(14)-H), 8.29 (1H, d, *J* = 7.8 Hz, N(5)-H). **^{13}C NMR** (125 MHz, methanol-d_3): δ 17.6, 18.9, 21.5, 23.5, (2 × C(3)-H$_3$/ 2 × C(18)-H$_3$), 25.77 (C(17)-H), 25.84 (C(22)-H$_2$), 30.7 (C(21)-H$_2$), 31.4 (C(2)-H), 39.8 (C(12)-H$_2$), 40.70, 40.71 (C(7)-H$_2$/C(16)-H$_2$), 48.4 (C(23)-H$_2$, assigned by HSQC), 50.8 (C(6)-H), 50.9 (C(15)-H), 60.0 (C(1)-H), 61.4 (C(20)-H), 127.6 (C(11)-H), 129.4, 130.3 (C(9)-H/C(10)-H), 139.1 (C(8)), 169.0 (C(4)=O), 172.8 (C(13)=O), 173.7 (C(19)=O), 176.8 (C(24)=O).

48

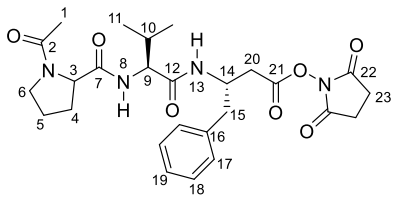


^{1}H NMR (500 MHz, acetonitrile-d_3): δ 0.88-0.95 (12H, m, 2 × C(3)-H$_3$, 2 × C(18)-H$_3$), 1.45 (2H, dd, C(15)-H$_2$), 1.55-1.69 (1H, m, C(16)-H), 1.90-1.99 (3H, m, C(20)-H, C(21)-H$_2$, assigned by TOCSY), 2.05-2.31 (2H, m, C(2)-H, C(21)-H', assigned by TOCSY), 2.92 (1H, dd, *J* = 8.0, 13.9 Hz, C(7)-H), 3.05 (1H, dd, *J* = 6.3, 13.9 Hz, C(7)-H'), 3.49-3.63 (2H, m, C(23)-H$_2$), 3.73 (1H, d, *J* = 5.2 Hz, C(1)-H), 4.27 (1H, dd, *J* = 4.0, 8.0 Hz, C(19)-H), 4.63-4.74 (2H, m, C(6)-H, C(14)-H), 5.68 (1H, s, N(24)-H), 6.42 (1H, s, N(24)-H'), 7.14 (1H, d, *J* = 6.5 Hz, N(13)-H).7.16-7.32 (5H, m, 2 × C(9)-H, 2 × C(10)-H, C(11)-H), 7.51 (1H, d, *J* = 8.1 Hz, N(5)-H). **^{13}C NMR** (125 MHz, acetonitrile-d_3): δ 17.5, 18.5 (2 × C(3)-H$_3$), 21.8, 23.6 (2 × C(17)-H$_3$), 25.3 (C(16)-H), 25.6 (C(21)-H$_2$), 29.6 (C(20)-H$_2$), 31.0 (C(2)-H),

[xi] NMR characterization of compounds in this chapter has been previously reported by Danny Burke.

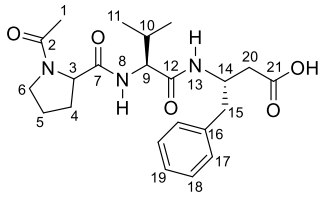
38.5 (C(7)-H_2), 41.8 (C(15)-H_2), 48.0 (C(22)-H_2), 50.0 (C(14)-H), 50.4 (C(6)-H), 59.6 (C(1)-H), 60.6 (C(19)-H), 127.6 (C(11)-H), 129.2, 130.3 (C(9)-H/C(10)-H), 137.8 (C(8)), 168.8 (C(4)=O), 171.0 (C(12)=O), 172.0 (C(18)=O), 174.6 (C(23)=O).

49



^{1}H NMR (500 MHz, acetonitrile-d_3): δ 0.66 (3H, d, *J* = 6.9 Hz, C(11)-H_3), 0.77 (3H, d, *J* = 6.9 Hz, C(11)-H_3'), 1.90-1.98 (2H, m, C(5)-H_2, assigned by TOCSY), 1.99-2.09 (6H, m, C(1)-H_3, C(4)-H_2, C(10)-H), 2.74-2.81 (5H, m, C(20)-H', 2 × C(23)-H_2), 2.81-2.97 (5H, m, C(15)-H_2, C(20)-H), 3.45-3.53 (1H, m, C(6)-H), 3.58-3.65 (1H, m, C(6)-H'), 4.08 (1H, dd, *J* = 5.5, 8.5 Hz, 1H, C(9)-H), 4.32-4.44 (2H, m, C(3)-H, C(14)-H), 6.91-7.02 (2H, m, N(8)-H, N(13)-H), 7.19-7.33 (5H, m, 2 × C(17)-H, 2 × C(18)-H, C(19)-H). **^{13}C NMR** (125 MHz, acetonitrile-d_3): δ 17.7, 19.7 (C(11)-H_3/C(11)-H_3'), 22.8 (C(1)-H_3), 25.6 (C(5)-H_2), 26.4 (2 × C(23)-H_2), 29.5 (C(4)-H_2), 30.8 (C(10)-H), 36.4 (C(20)-H_2), 40.0 (C(15)-H_2), 48.9 (C(14)-H), 49.2 (C(6)-H_2), 59.2 (C(9)-H), 61.4 (C(3)-H), 127.6 (C(19)-H), 129.4 (C(18)-H), 130.3 (C(17)-H), 139.0 (C(16)), 167.8 (2 × C(21)=O), 171.1, 171.8 (C(12)=O/C(22)=O), 172.2 (C(2)=O), 172.8 (C(7)=O).

50



^{1}H NMR (500 MHz, acetonitrile-d_3): δ 0.69 (3H, d, *J* = 6.9 Hz, C(11)-H_3), 0.77 (3H, d, *J* = 6.9 Hz, C(11)-H_3'), 1.91-1.98 (2H, m, C(5)-H_2, assigned by TOCSY), 1.99-2.05 (3H, m, C(4)-H_2, C(10)-H), 2.07 (3H, s, C(1)-H_3), 2.30-2.47 (2H, m, C(20)-H_2, assigned by TOCSY), 2.76-2.86 (2H, m, C(15)-H_2), 3.45-3.52 (1H, m, C(6)-H), 3.62 (1H, ddd, *J* = 5.0, 6.7, 9.9 Hz, C(6)-H'), 4.00 (1H, dd, *J* = 5.7, 8.5 Hz, 1H, C(9)-H), 4.32-4.40 (2H, m, C(3)-H, C(14)-H), 6.87-6.94 (2H, m, N(8)-H, N(13)-H), 7.17-7.23 (3H, m, 2 × C(17)-H, C(19)-H), 7.25-7.31 (2H, m, 2 × C(18)-H). **^{13}C NMR** (125 MHz, acetonitrile-d_3): δ 16.9, 18.7 (C(11)-H_3/C(11)-H_3'), 21.9 (C(1)-H_3), 24.7 (C(5)-H_2), 28.6 (C(4)-H_2), 29.8 (C(10)-H), 38.3 (C(20)-H_2), 39.8 (C(15)-H_2), 47.8 (C(14)-H), 48.1 (C(6)-H_2), 58.5 (C(9)-H), 60.3 (C(3)-H), 126.3

(C(19)-H), 128.3 (C(18)-H), 129.3 (C(17)-H), 138.5 (C(16)), 170.4 (C(12)=O), 171.0 (C(2)=O), 171.78, 178.81 (C(7)=O/C(21)=O).

Chapter 8. Appendix

8.1. HPLC traces and MALDI-TOF MS for synthesised peptide sequences.

8.1.1. Chapter 2

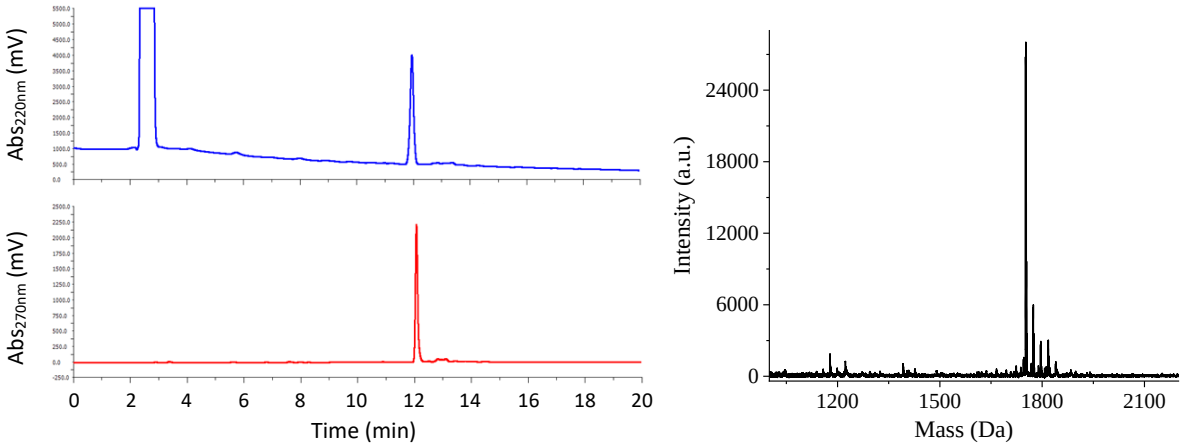


Figure 8.1: 18 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 12.1 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1752.182 Da, observed mass $[M+Na]^+$ = 1751.706 Da.

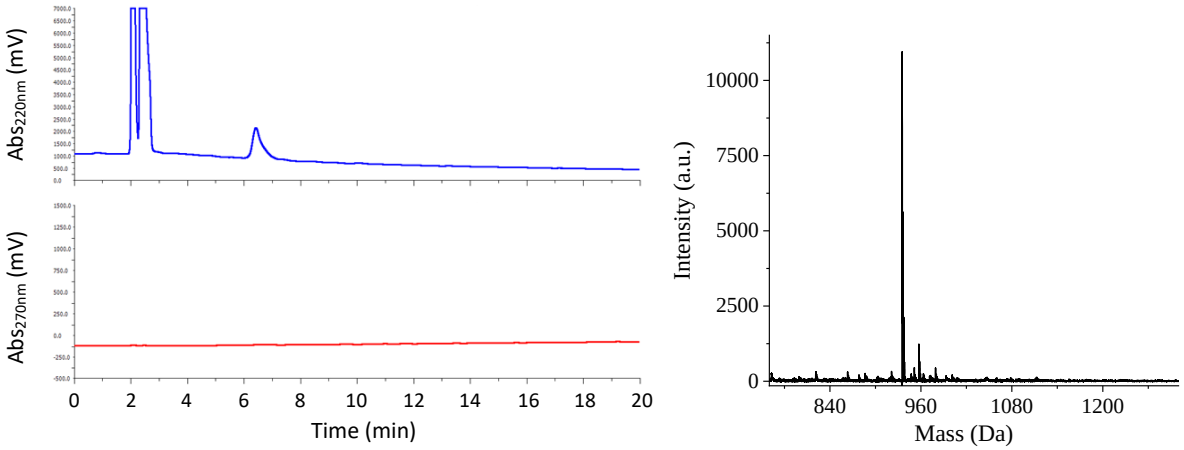


Figure 8.2: 19 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 6.4 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 935.358 Da, observed mass $[M-H]^-$ = 935.357 Da.

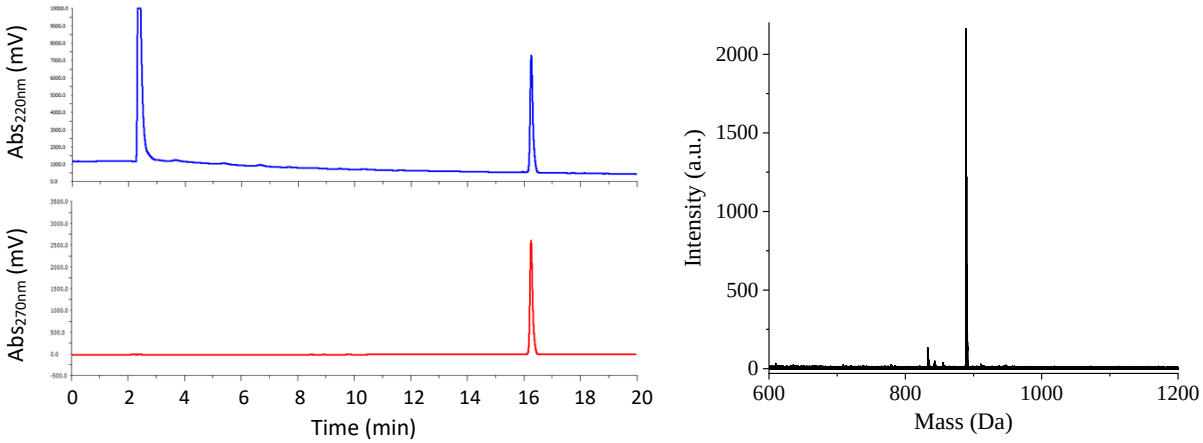


Figure 8.3: 21a - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 16.3 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 889.414 Da, observed mass $[M-H]^-$ = 888.862 Da.

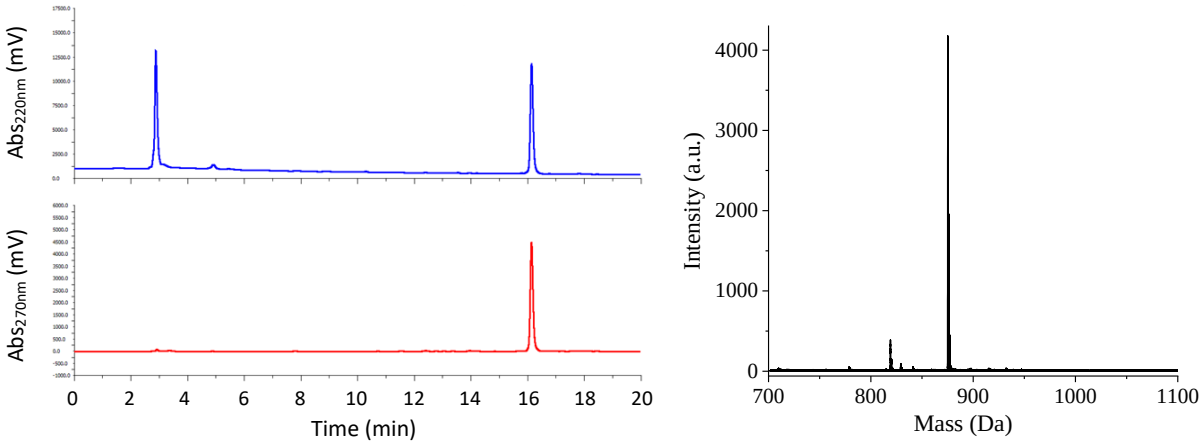


Figure 8.4: 21b - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 16.2 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 875.399 Da, observed mass $[M-H]^-$ = 875.217 Da.

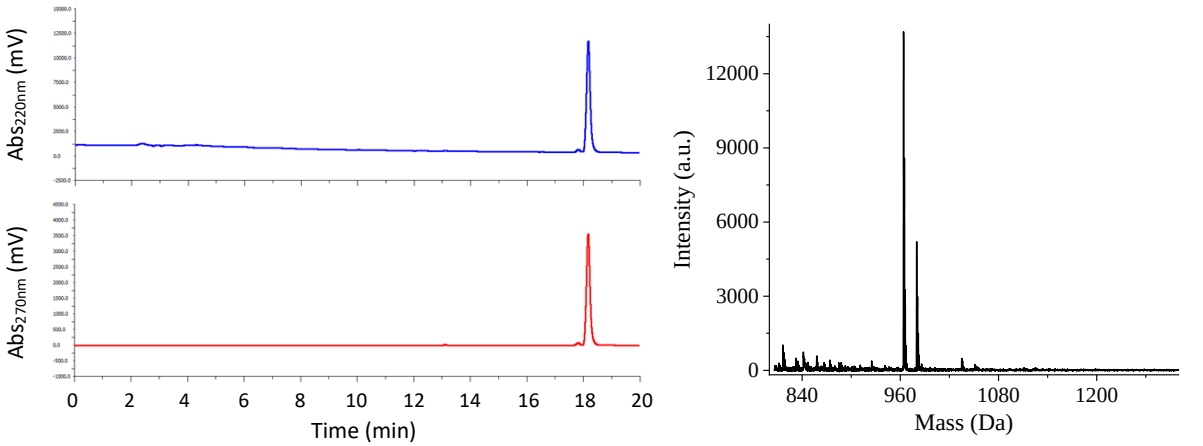


Figure 8.5: 22 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 18.9 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 963.372 Da, observed mass $[M+Na]^+$ = 964.011 Da.

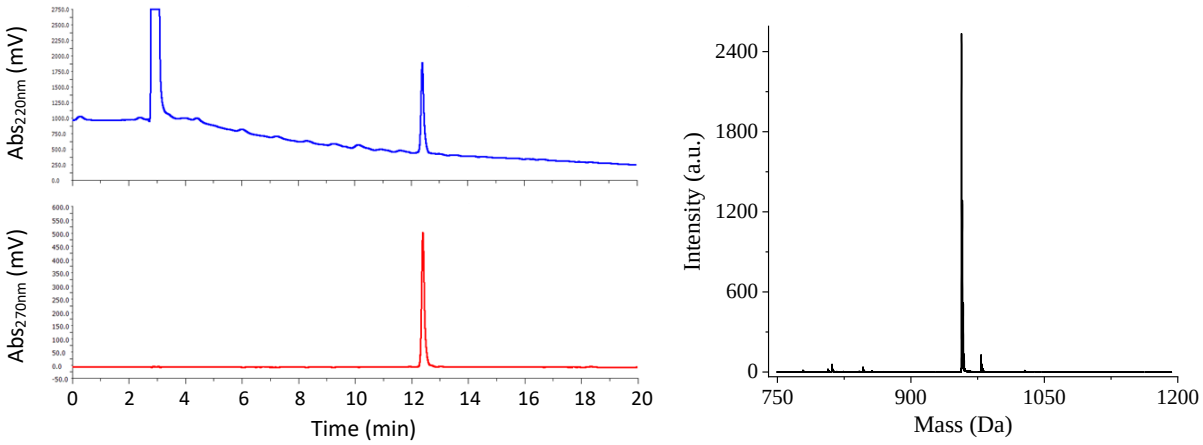


Figure 8.6: 23 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 12.4 min. (right) MALDI-TOF MS. Calculated mass $[M\text{-}H]^-$ = 956.933 Da, observed mass $[M\text{-}Na]^-$ = 957.317 Da.

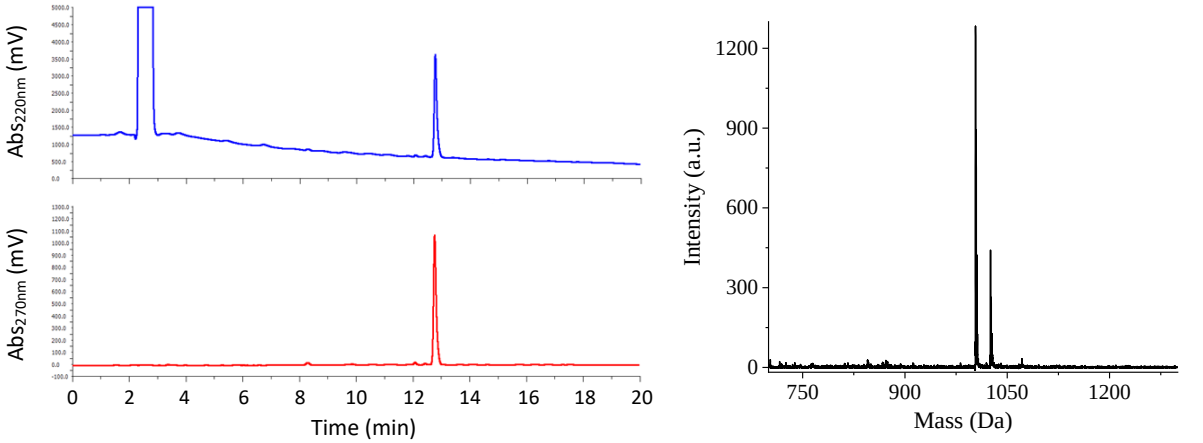


Figure 8.7: 23' - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3), t_R = 12.8 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1003.483 Da, observed mass $[M+Na]^+$ = 1003.295 Da.

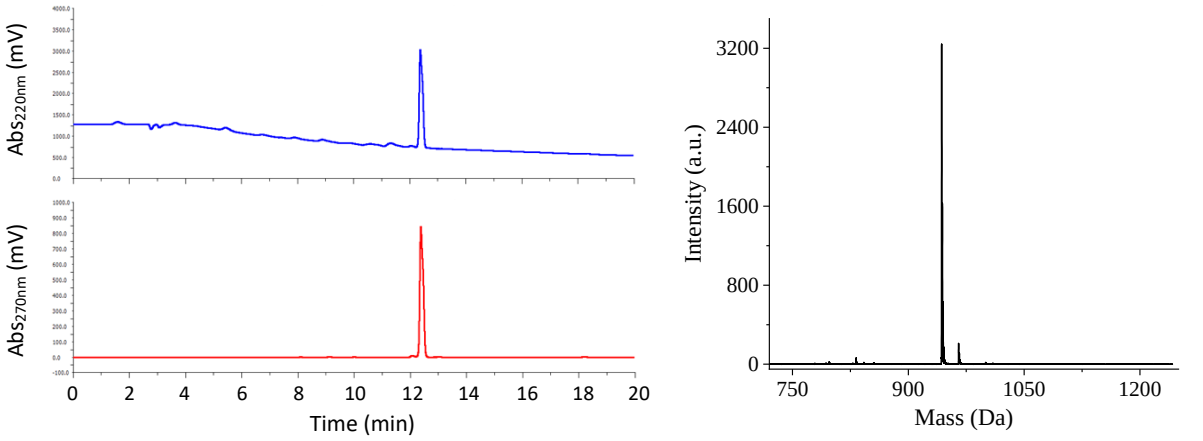


Figure 8.8: 24 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 12.4 min. (right) MALDI-TOF MS. Calculated mass $[M\text{-}H]^-$ = 943.128 Da, observed mass $[M\text{-}H]^-$ = 943.301 Da.

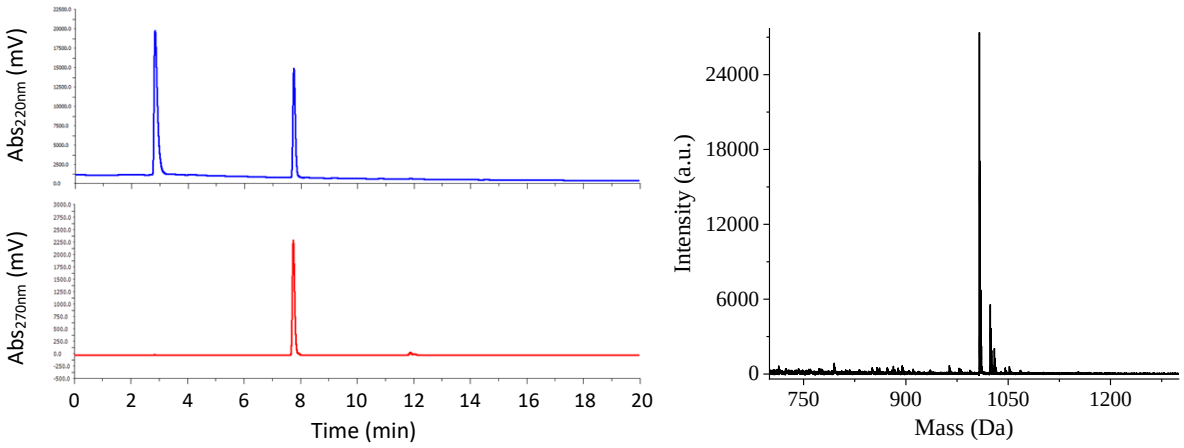


Figure 8.9: 25 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 7.8 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1007.362 Da, observed mass $[M+Na]^+$ = 1007.764 Da.

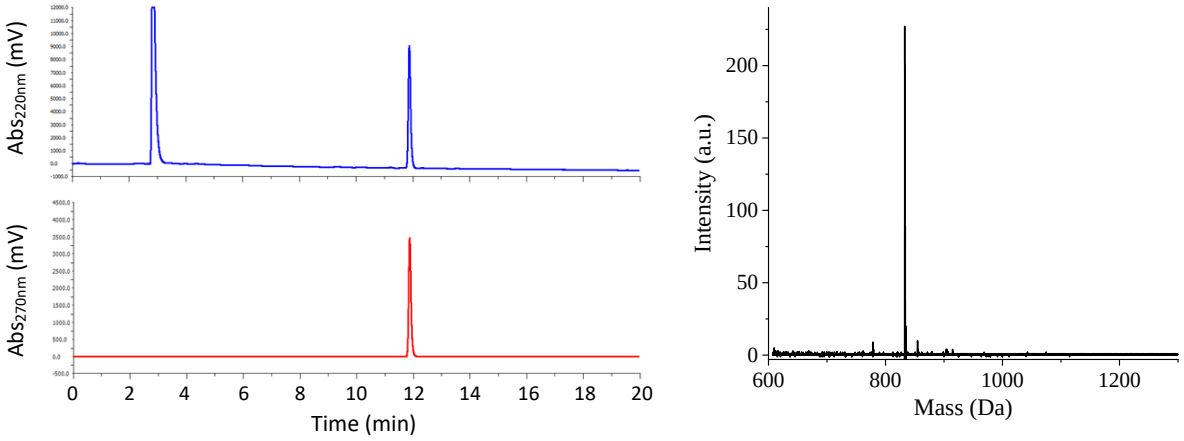


Figure 8.10: 26 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 11.9 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 833.352 Da, observed mass $[M-H]^-$ = 832.821 Da.

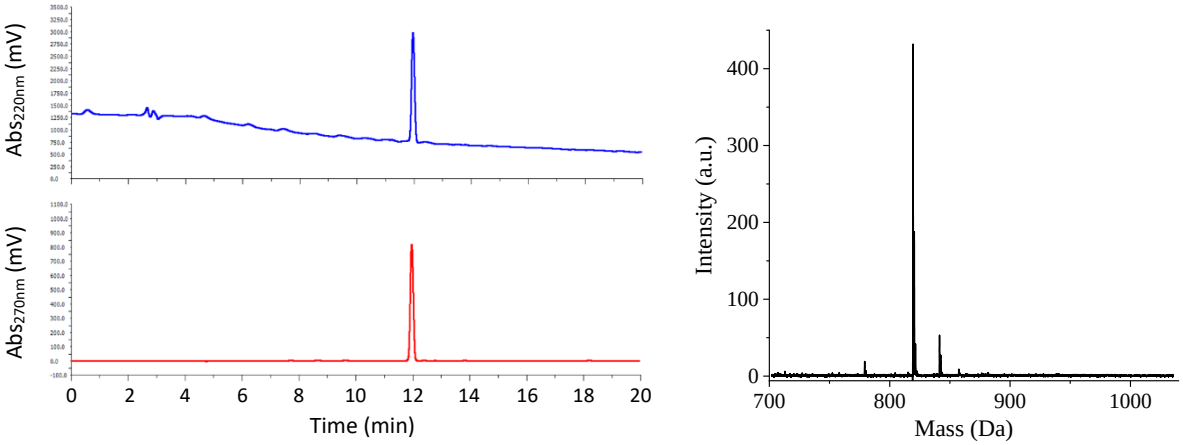


Figure 8.11: 27 - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 12.0 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 819.336 Da, observed mass $[M-H]^-$ 819.407 = Da.

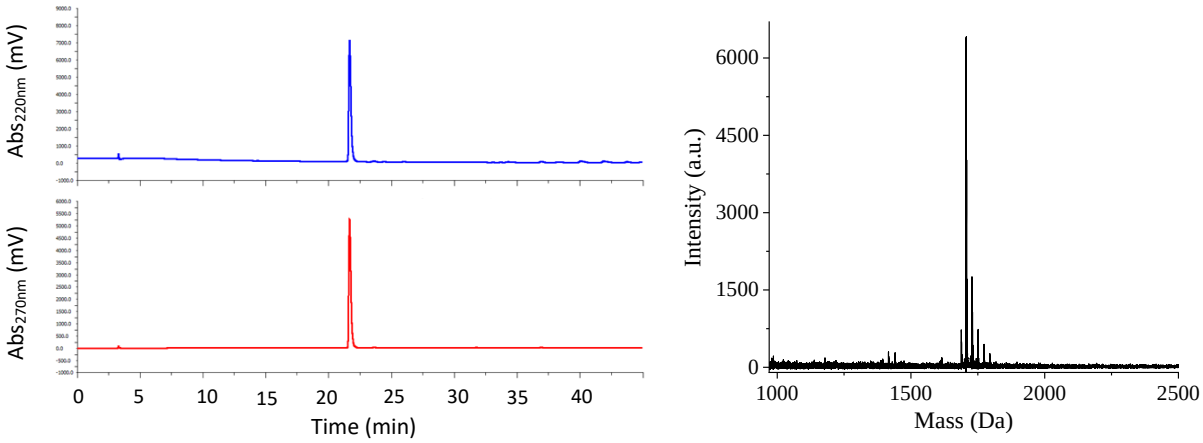


Figure 8.12: 1 - (left) HPLC traces, gradient: 0-100% ACN in H_2O (0.1 M NH_4HCO_3) over 45 min, t_R = 21.7 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1705.649 Da, observed mass $[M+Na]^+$ = 1706.248 Da.

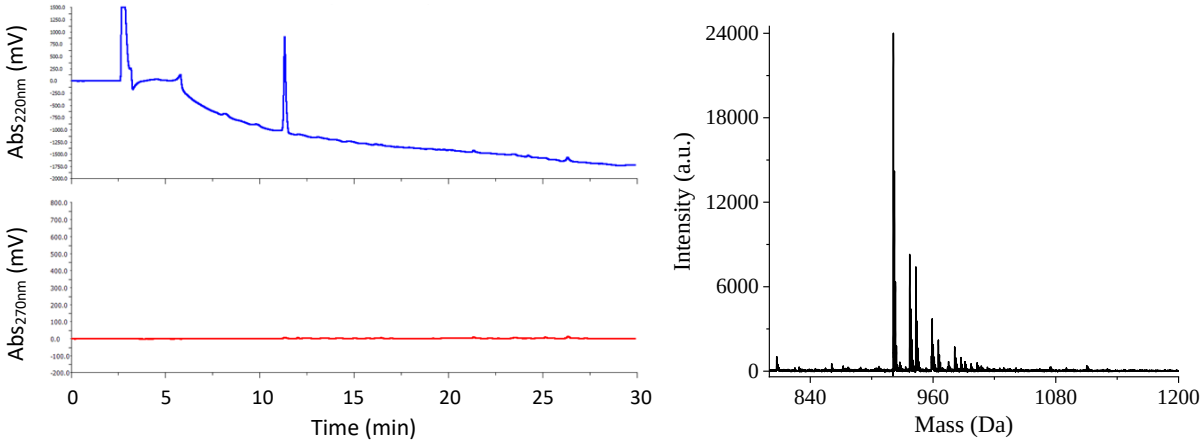


Figure 8.13: 28 - (left) HPLC traces, gradient: 0-100% ACN in H_2O (0.1 M NH_4HCO_3) over 30 min, t_R = 11.3 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 921.342 Da, observed mass $[M+Na]^+$ = 921.249 Da.

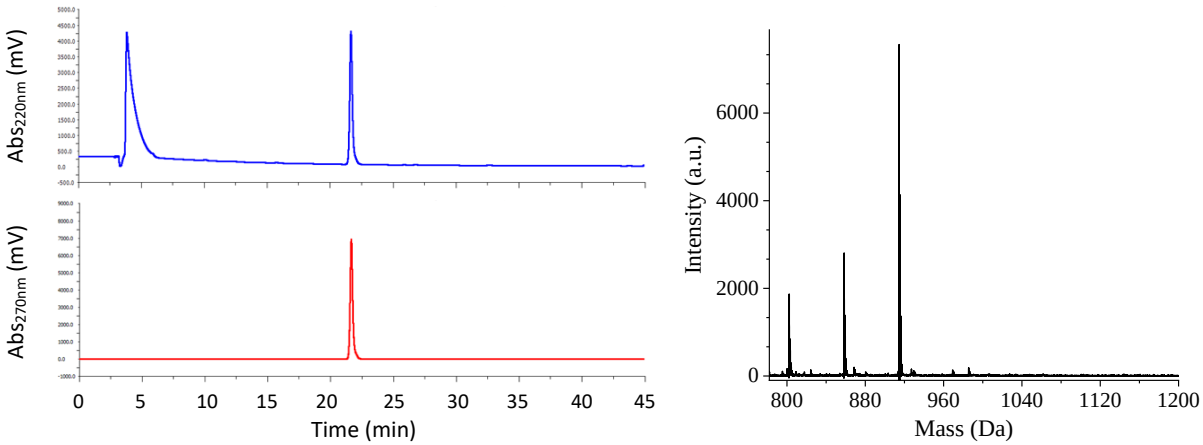


Figure 8.14: 29 - (left) HPLC traces, gradient: 0-100% ACN in H_2O (0.1 M NH_4HCO_3) over 45 min, t_R = 21.6 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 914.504 Da, observed mass $[M-H]^-$ = 913.435 Da.

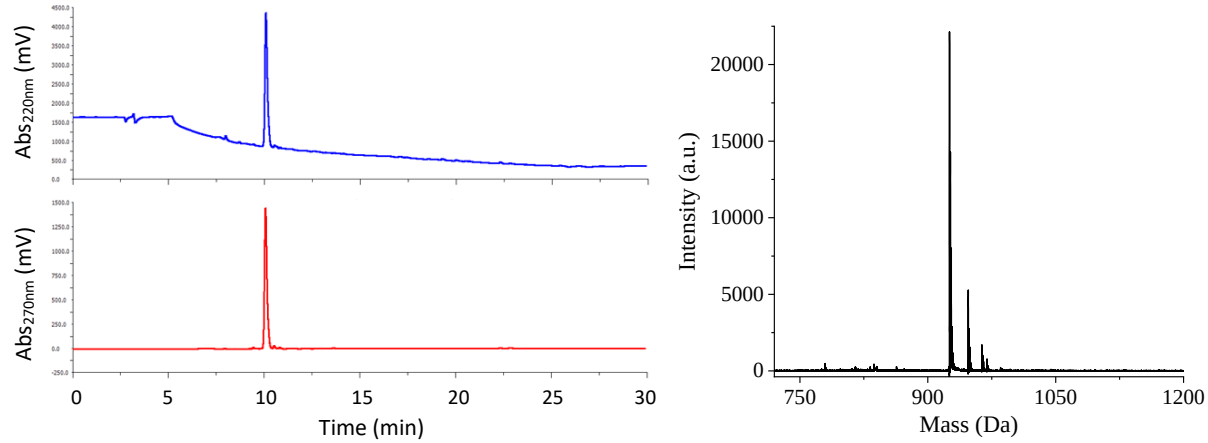


Figure 8.15: 30 - (left) HPLC traces, gradient: 0-100% ACN in H_2O (0.1 M NH_4HCO_3) over 30 min, t_R = 10.1 min. (right) MALDI-TOF MS. Calculated mass $[M-H]^-$ = 925.485 Da, observed mass $[M-H]^-$ = 925.275 Da.

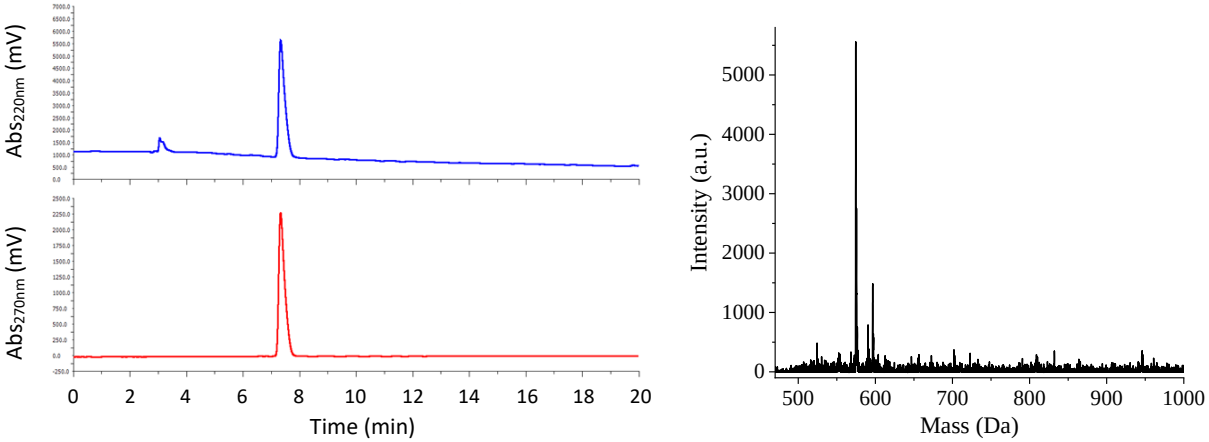


Figure 8.16: Std - (left) HPLC traces, gradient: 10-60% ACN in H_2O (0.1 M NH_4HCO_3) over 20 min, t_R = 7.3 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 574.227 Da, observed mass $[M+Na]^+$ = 574.482 Da.

8.1.2. Chapter 3

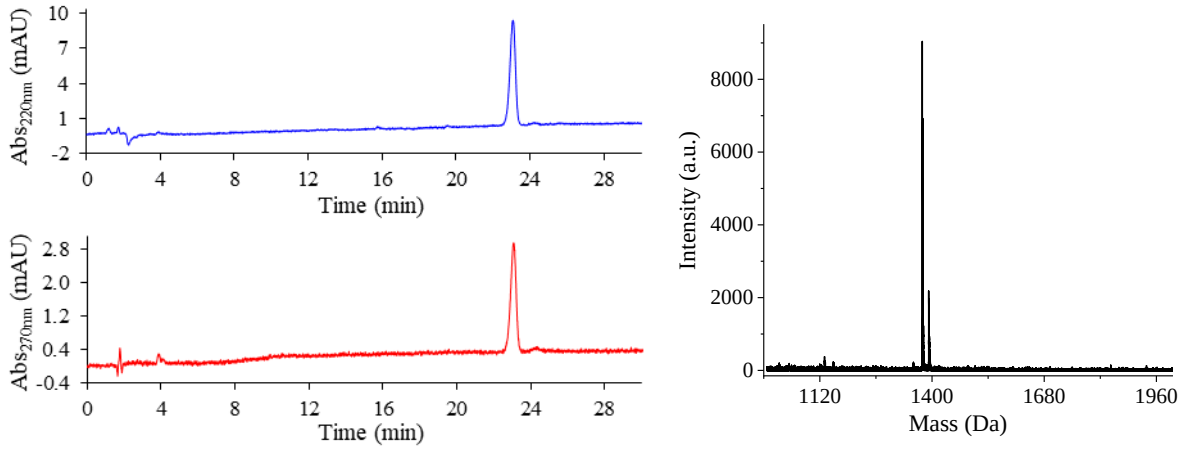


Figure 8.17: 34 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 30 min, t_R = 23.1 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1375.768 Da, observed mass $[M+Na]^+$ = 1375.740 Da.

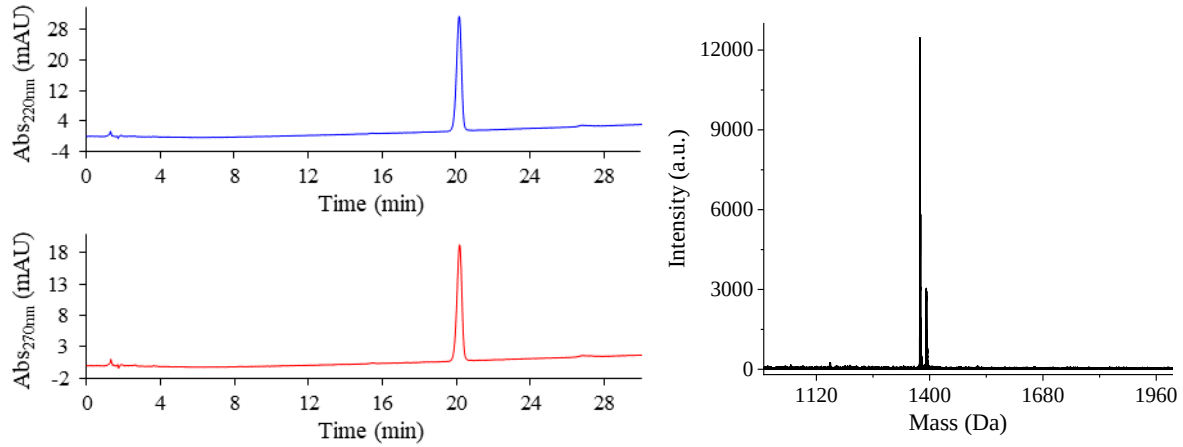


Figure 8.18: 35 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 30 min, t_R = 20.2 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1375.768 Da, observed mass $[M+Na]^+$ = 1375.654 Da.

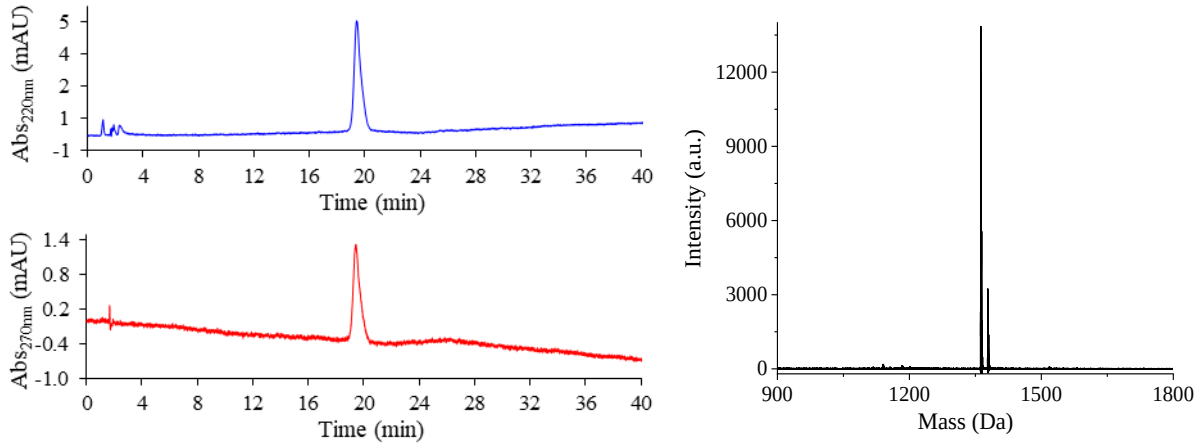


Figure 8.19: 36 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 40 min, t_R = 19.5 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1361.752 Da, observed mass $[M+Na]^+$ = 1362.288 Da.

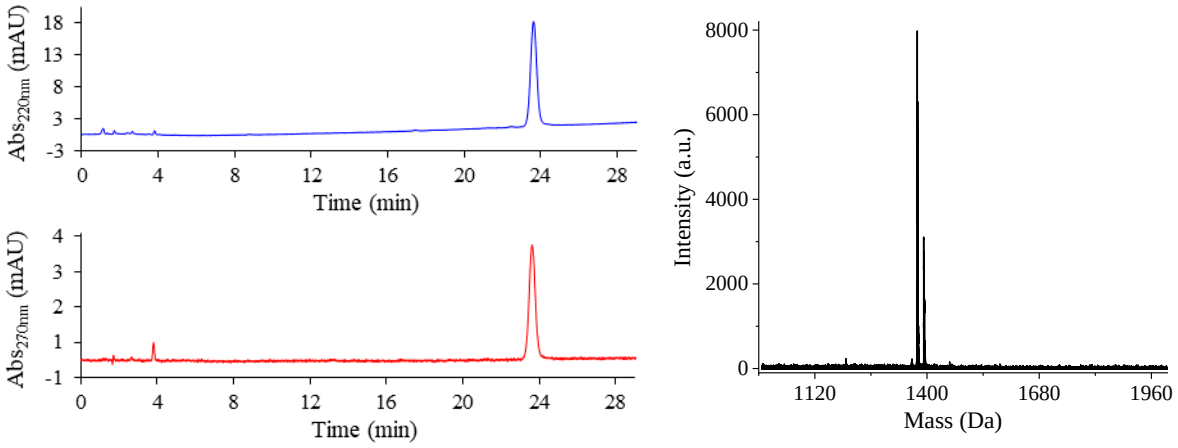


Figure 8.20: 38 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 30 min, t_R = 23.7 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1375.768 Da, observed mass $[M+Na]^+$ = 1375.721 Da.

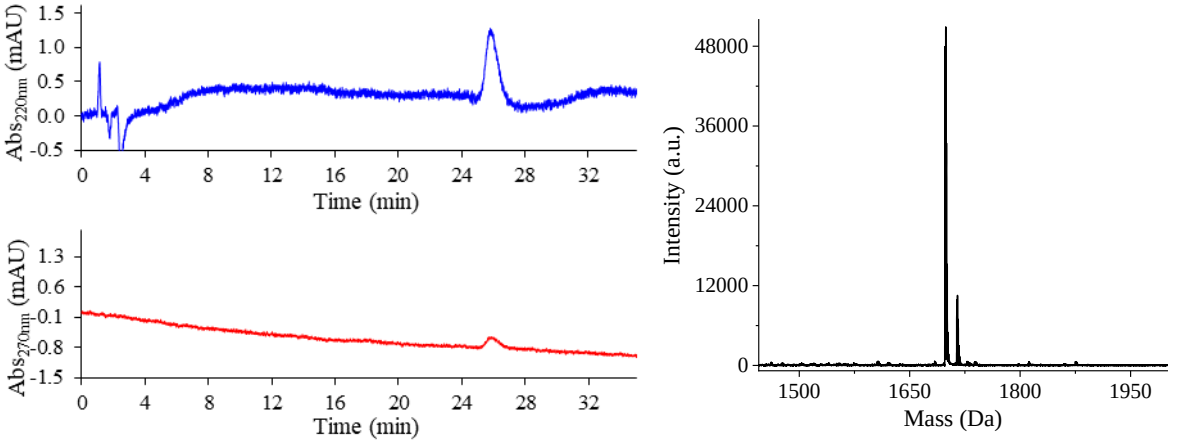


Figure 8.21: 39 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 35 min, t_R = 25.9 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1697.935 Da, observed mass $[M+Na]^+$ = 1697.947 Da.

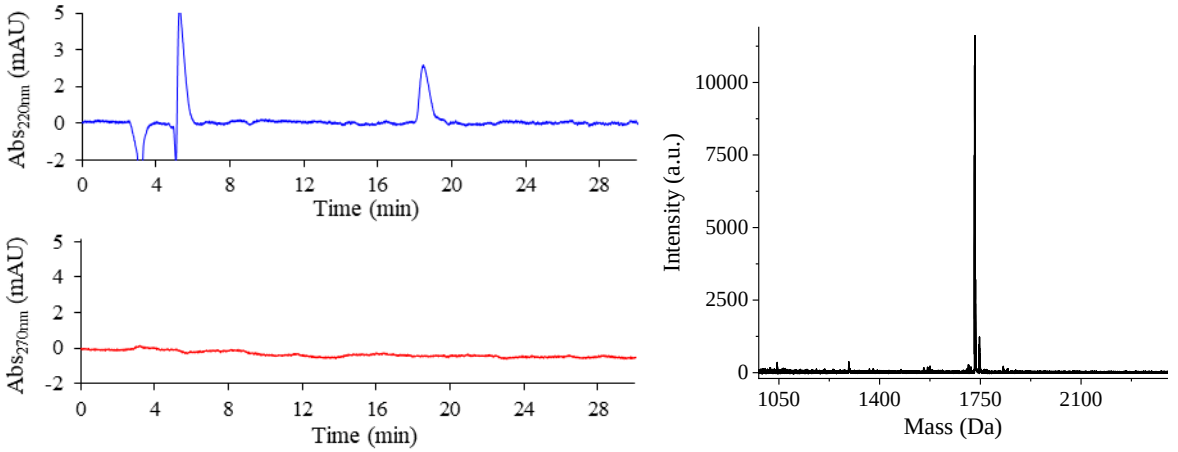


Figure 8.22: 40 - (left) HPLC traces, gradient: 50-100% MeOH in H_2O over 30 min, t_R = 18.6 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1730.031 Da, observed mass $[M+Na]^+$ = 1730.150 Da.

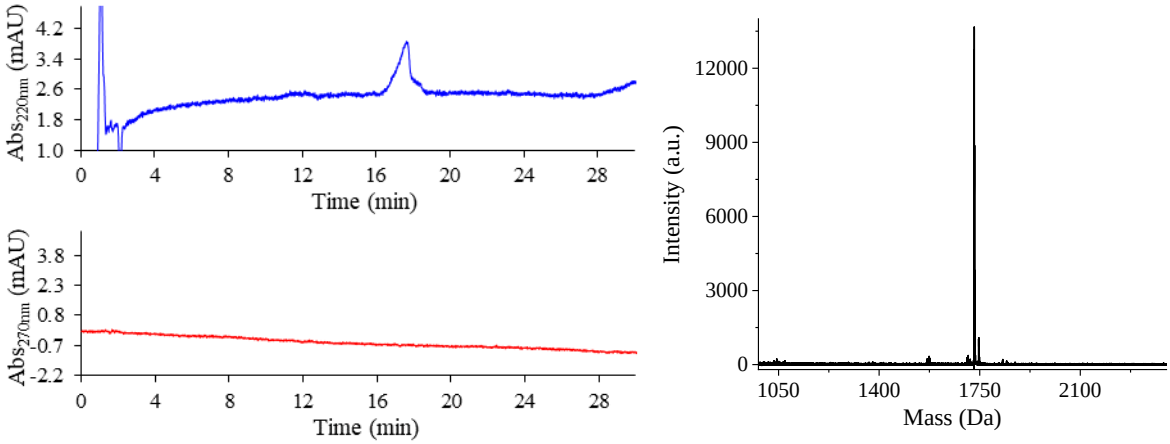


Figure 8.23: 41 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 30 min, t_R = 17.7 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1730.031 Da, observed mass $[M+Na]^+$ = 1730.193 Da.

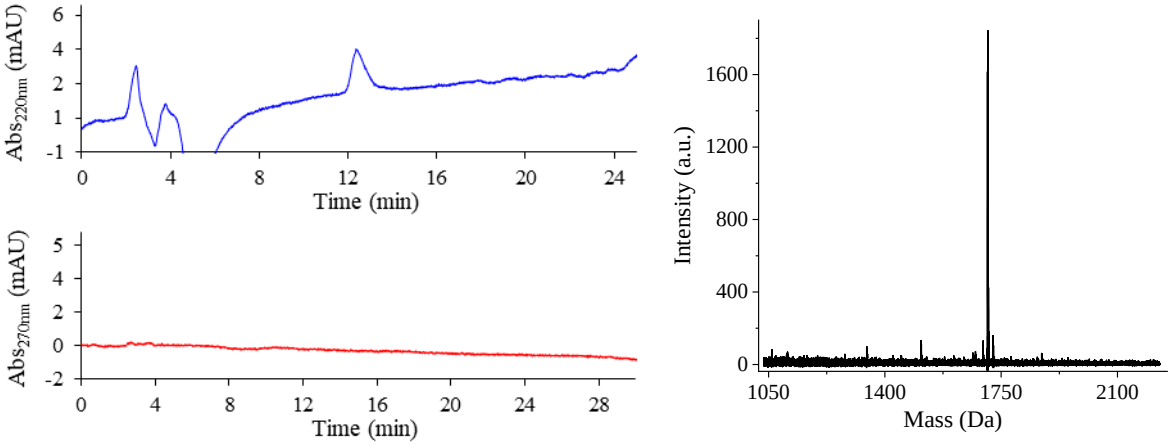


Figure 8.24: 42 - (left) HPLC traces, gradient: 70-100% MeOH in H_2O over 25 min, t_R = 12.6 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1708.933 Da, observed mass $[M+Na]^+$ = 1709.862 Da.

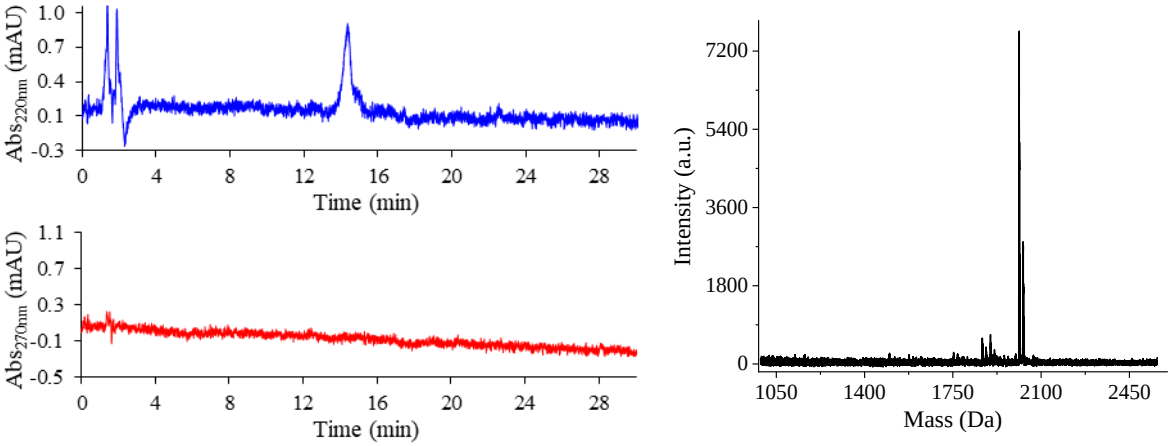


Figure 8.25: 43 - (left) HPLC traces, gradient: 80-100% MeOH in H_2O over 30 min, t_R = 14.4 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 2010.183 Da, observed mass $[M+Na]^+$ = 2010.961 Da.

8.1.3. Chapter 4[xii]

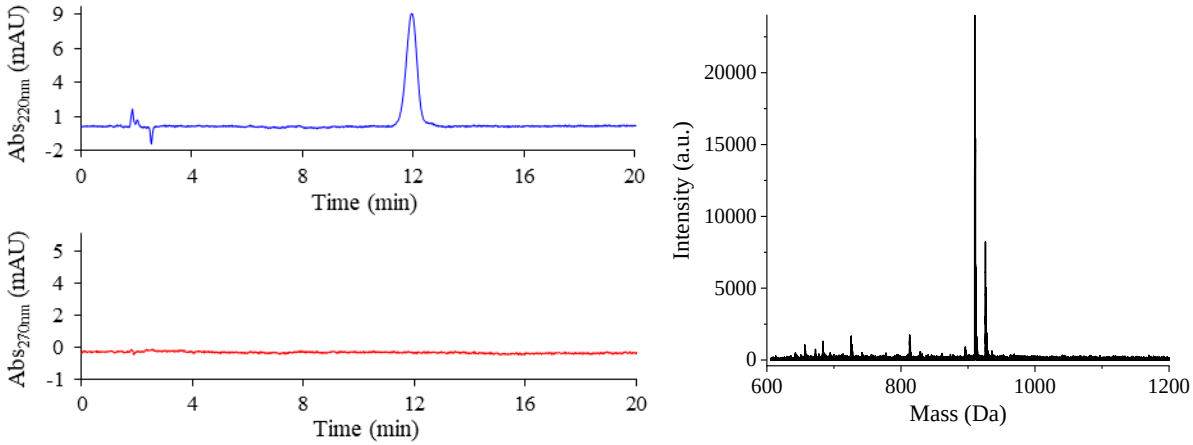


Figure 8.26: 45 - (left) HPLC traces, gradient: 70-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 20 min, t_R = 11.9 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 910.051 Da, observed mass $[M+Na]^+$ = 909.521 Da.

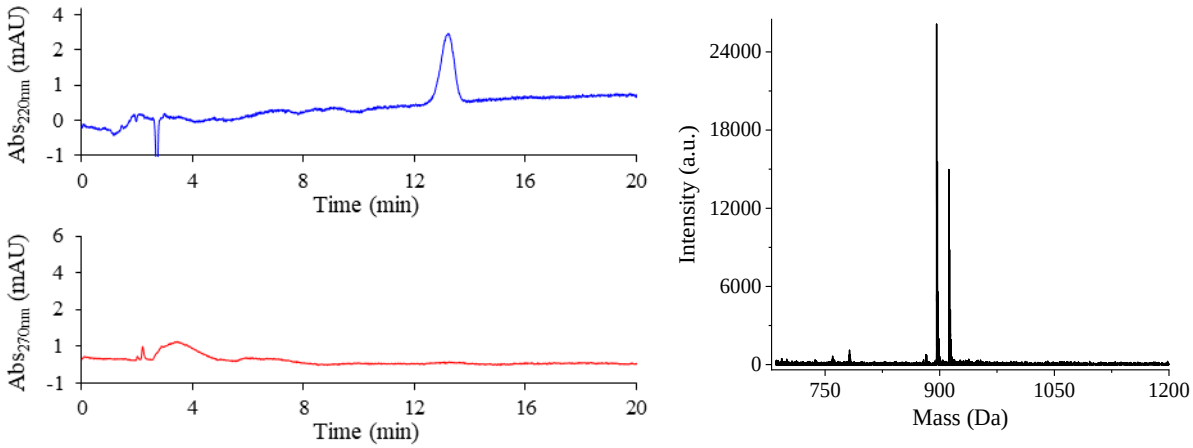


Figure 8.27: 46 - (left) HPLC traces, gradient: 70-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 20 min, t_R = 13.3 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 895.505 Da, observed mass $[M+Na]^+$ = 895.939 Da.

[xii] Analytical HPLC traces of compounds **45**, **46**, **47** and **48** were acquired by Danny Burke.

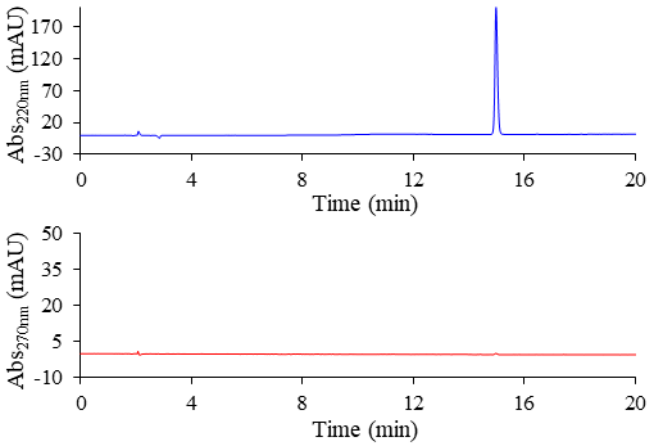


Figure 8.28: 47 - HPLC traces, gradient: 10-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 20 min, t_R = 15.0 min. ESI-HRMS.[xiii] Calculated mass $[M+H]^+$ = 488.323 Da, observed mass $[M+H]^+$ = 488.323 Da.

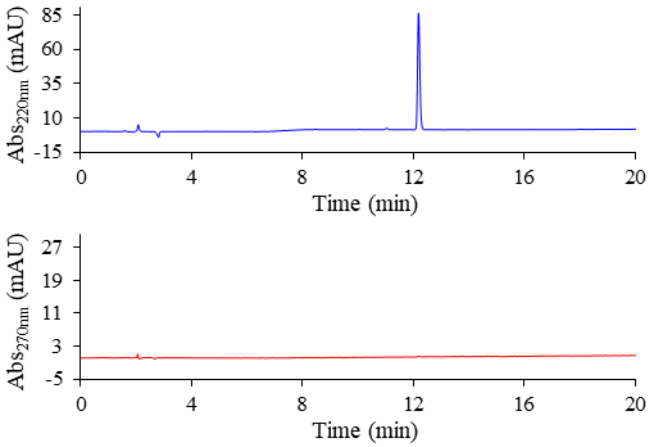


Figure 8.29: 48 - HPLC traces, gradient: 10-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 20 min, t_R = 12.2 min. ESI-HRMS. Calculated mass $[M+H]^+$ = 474.308 Da, observed mass $[M+H]^+$ = 474.309 Da.

[xiii] Compounds **47**-**50** could not be identified by MALDI-TOF MS due to their low molecular weight and the interference it had with the Matrix used in this technique. Instead high resolution ESI MS was acquired through the University of Bristol Mass Spectrometry Service.

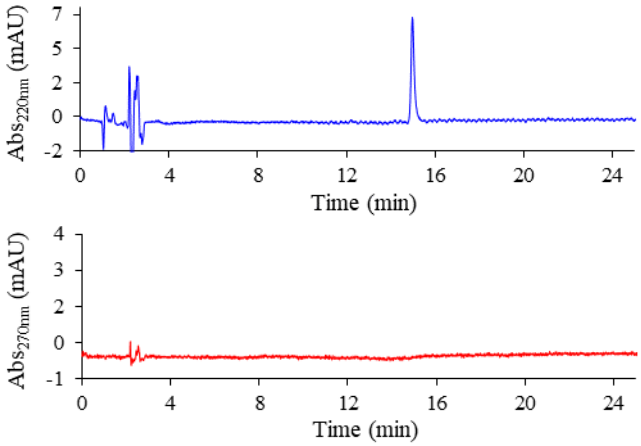


Figure 8.30: 50 - HPLC traces, gradient: 10-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 25 min, t_R = 15.0 min. ESI-HRMS. Calculated mass $[M+H]^+$ = 418.234 Da, observed mass $[M+H]^+$ = 418.232 Da.

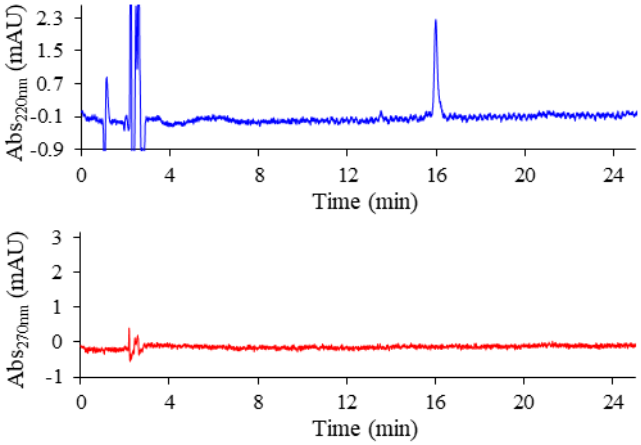


Figure 8.31: 49 - HPLC traces, gradient: 10-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 25 min, t_R = 16.0 min. (right) ESI-HRMS. Calculated mass $[M+H]^+$ = 515.250 Da, observed mass $[M+H]^+$ = 515.248 Da.

8.1.4. Chapter 5

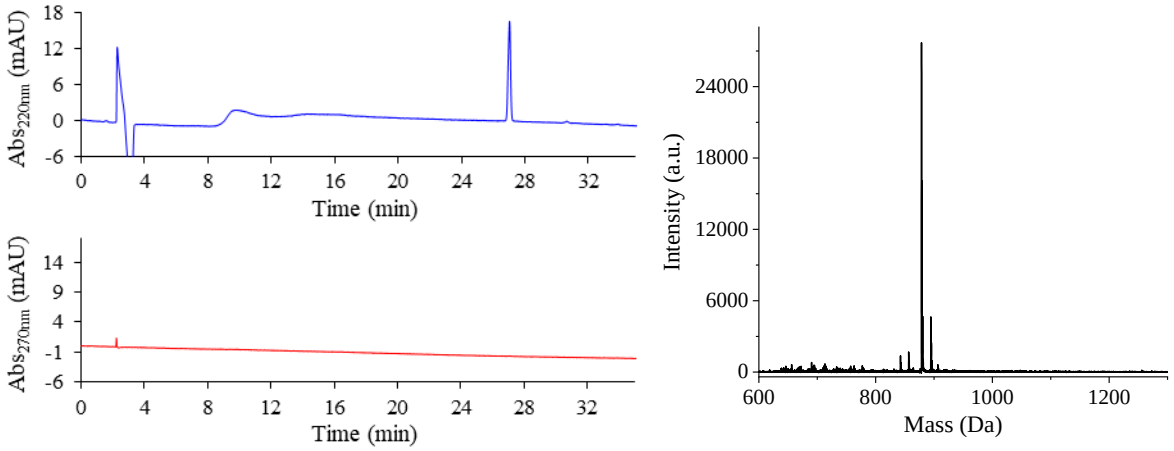


Figure 8.32: 51 - (left) HPLC traces, gradient: 0-100% ACN (0.1% TFA) in H_2O (0.1% TFA) over 35 min, t_R = 27.0 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 878.547 Da, observed mass $[M+Na]^+$ = 878.592 Da.

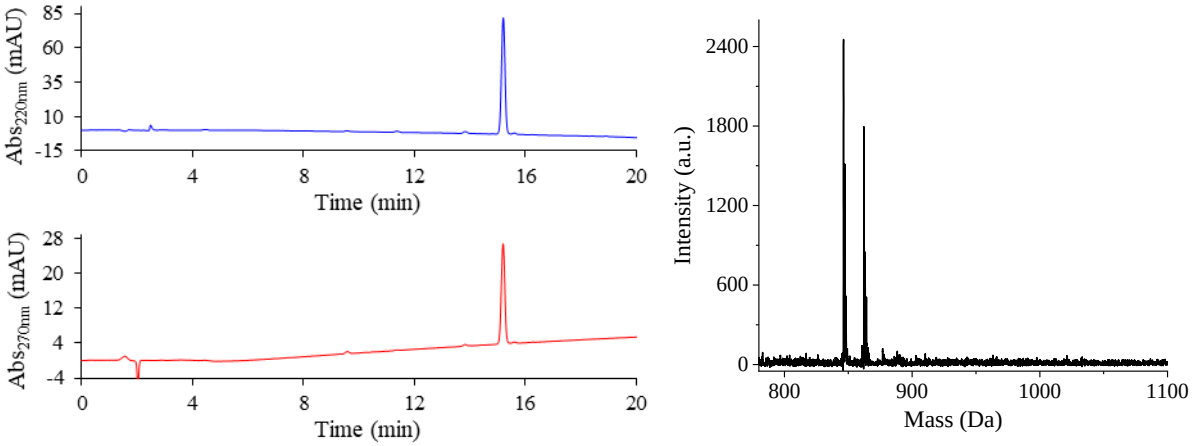


Figure 8.33: 52 - (left) HPLC traces, gradient: 50-100% MeOH in H_2O over 20 min, t_R = 15.2 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 846.405 Da, observed mass $[M+Na]^+$ = 846.285 Da.

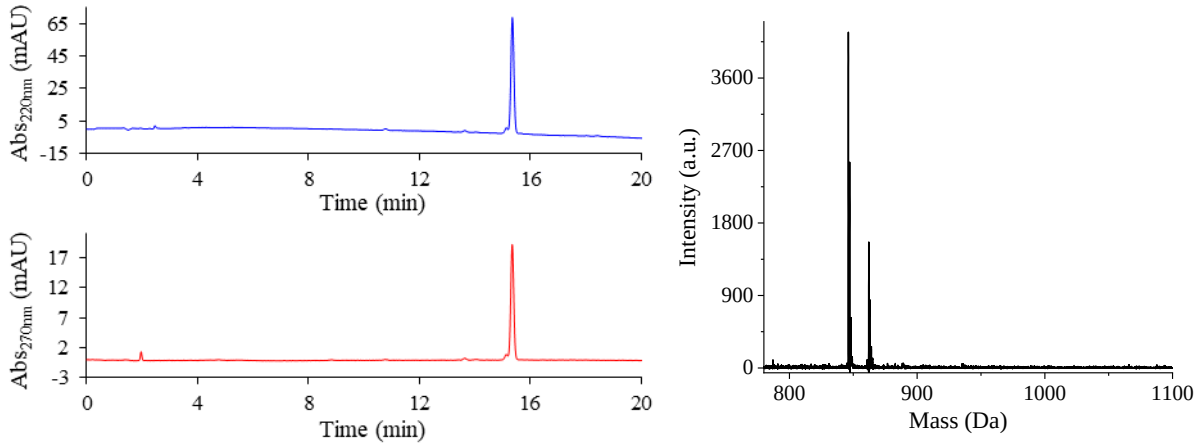


Figure 8.34: 53 - (left) HPLC traces, gradient: 50-100% MeOH in H_2O over 20 min, t_R = 15.3 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 846.405 Da, observed mass $[M+Na]^+$ = 846.166 Da.

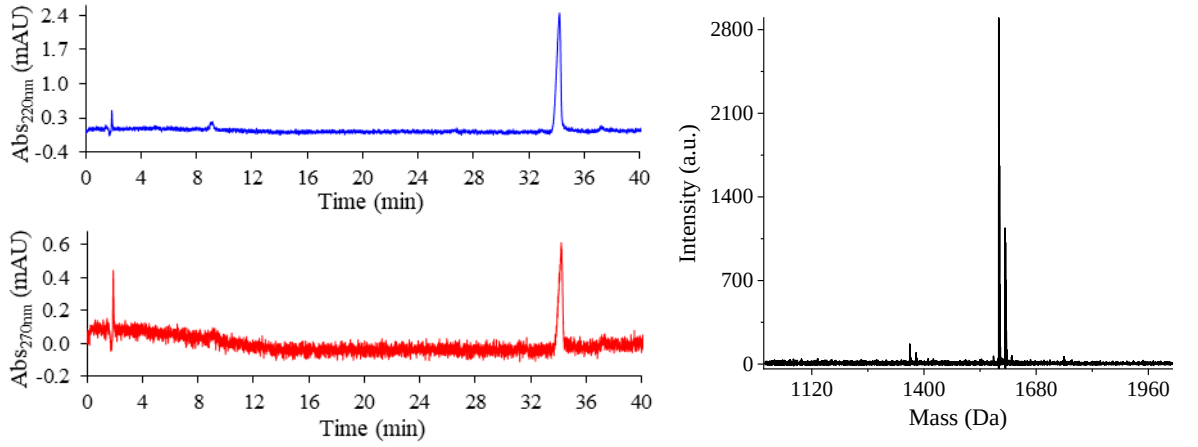


Figure 8.35: 54 - (left) HPLC traces, gradient: 70-100% MeOH in H_2O over 40 min, t_R = 34.2 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1586.936 Da, observed mass $[M+Na]^+$ = 1586.538 Da.

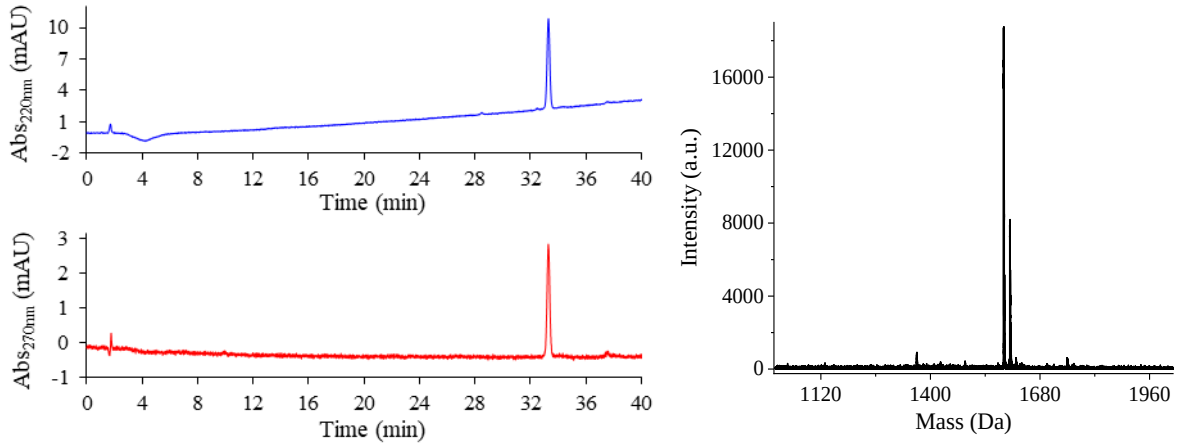


Figure 8.36: 55 - (left) HPLC traces, gradient: 70-100% MeOH in H_2O over 40 min, t_R = 33.3 min. (right) MALDI-TOF MS. Calculated mass $[M+Na]^+$ = 1586.936 Da, observed mass $[M+Na]^+$ = 1586.546 Da.

8.2. Circular dichroism data of selected peptide sequences.

8.2.1. Chapter3

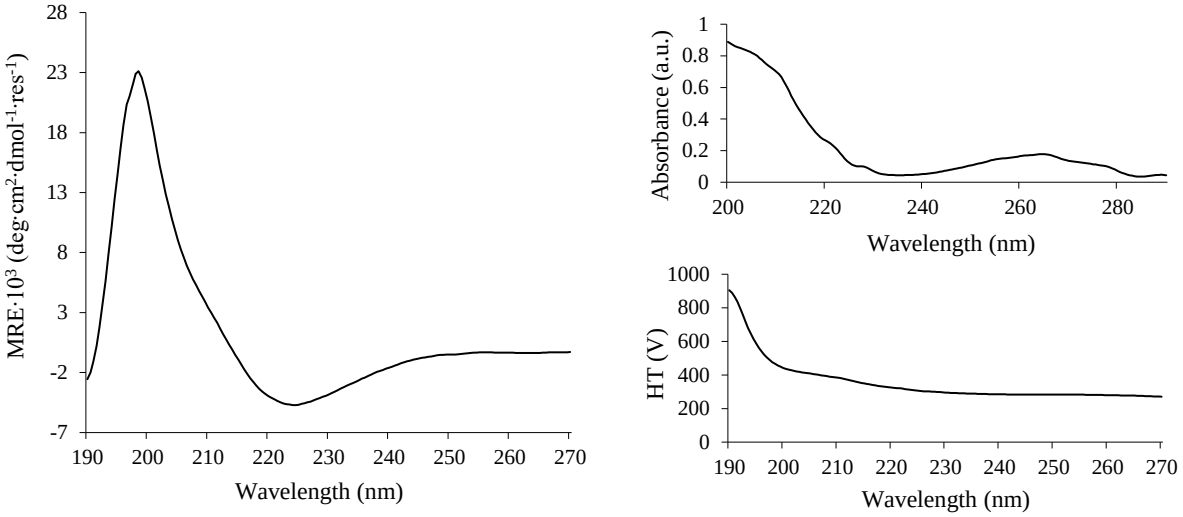


Figure 8.37: 34 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 101 µM peptide concentration in MeOH.

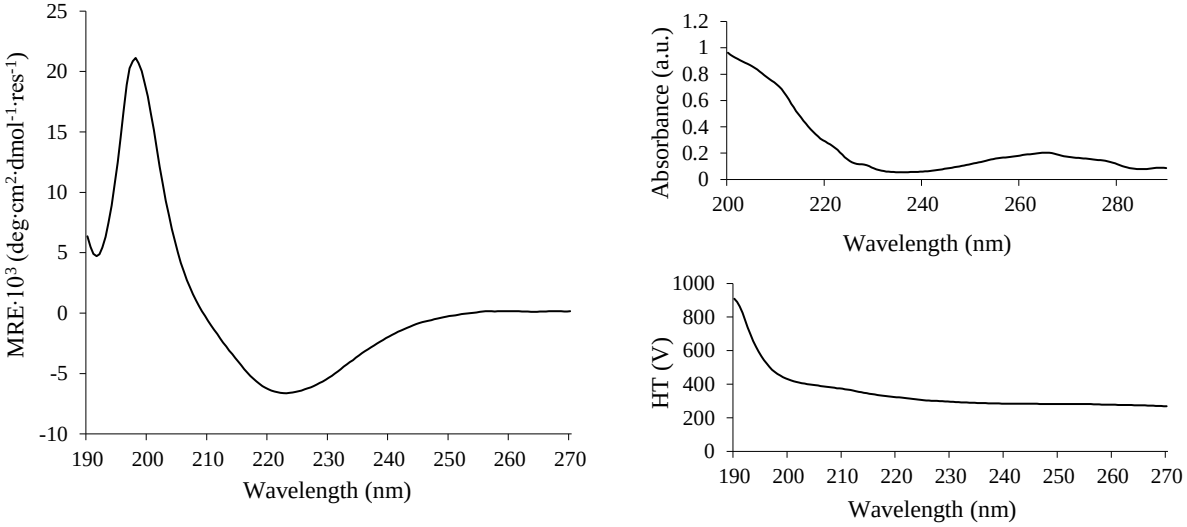


Figure 8.38: 35 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 112 µM peptide concentration in MeOH.

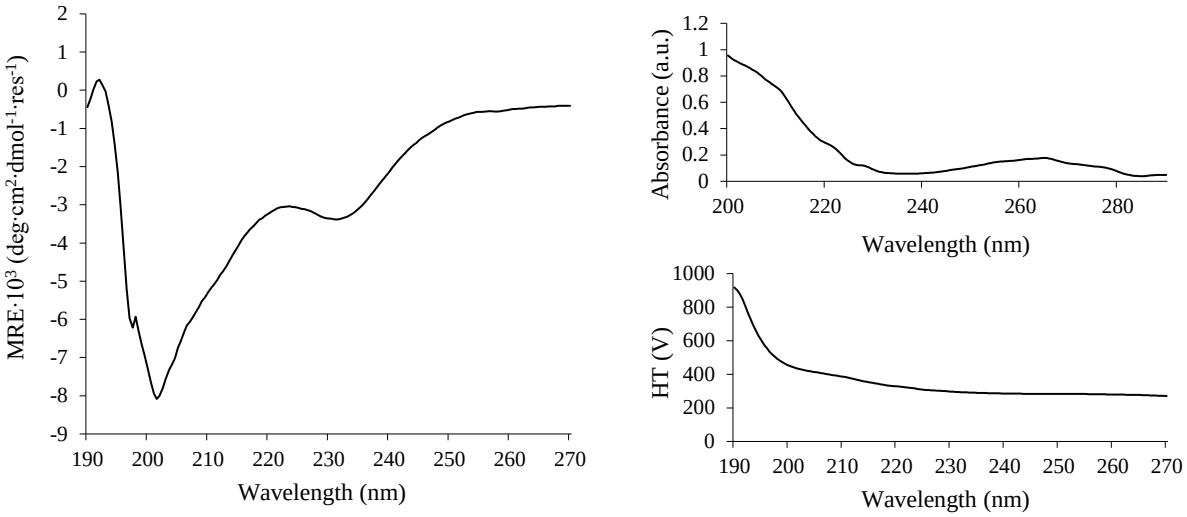


Figure 8.39: 36 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 101 µM peptide concentration in MeOH.

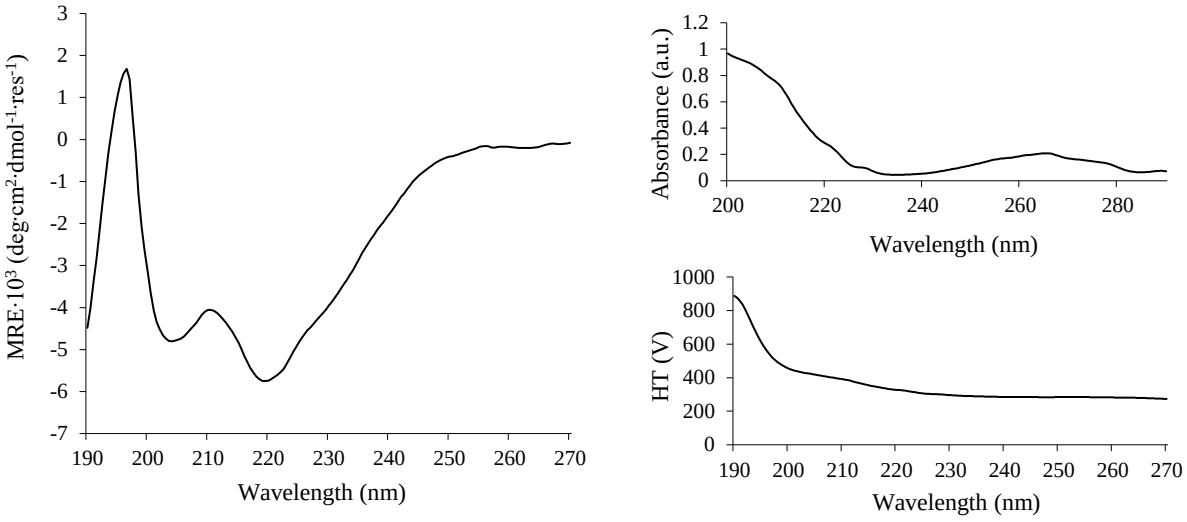


Figure 8.40: 38 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 114 µM peptide concentration in MeOH.

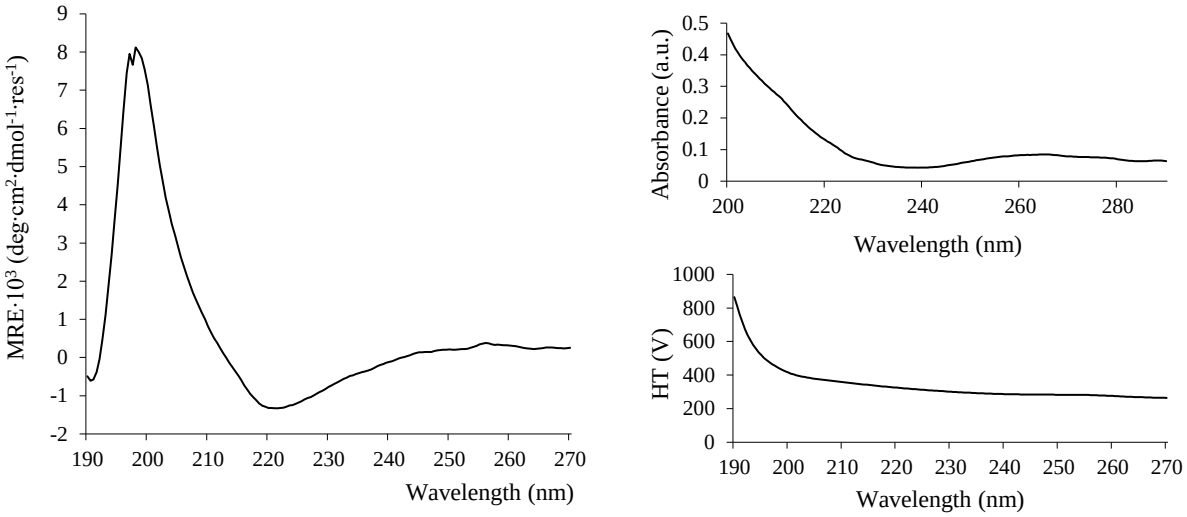


Figure 8.41: 39 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 47 µM peptide concentration in MeOH.

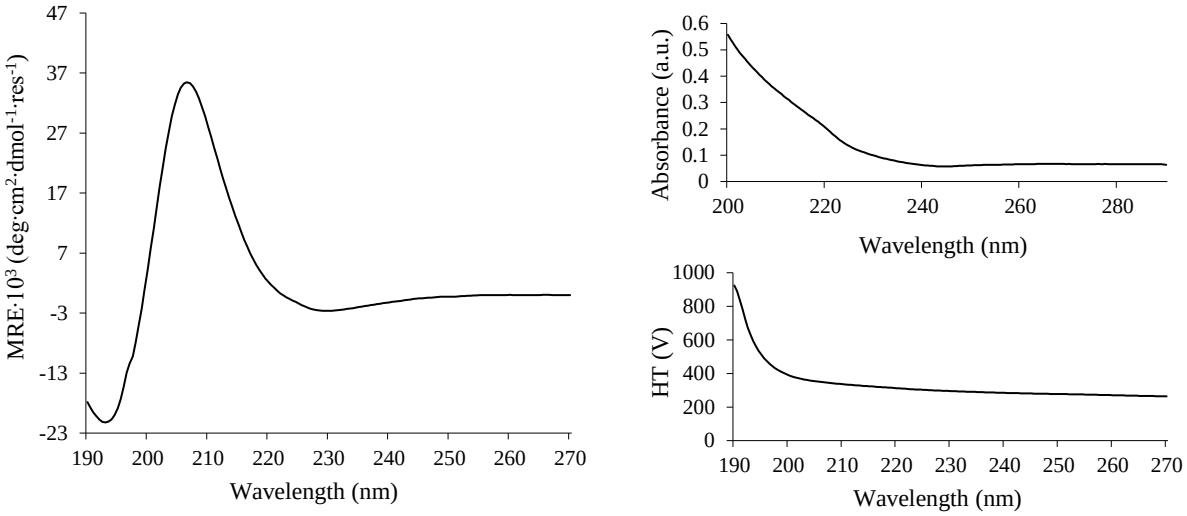


Figure 8.42: 40 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 8 µM peptide concentration in MeOH.

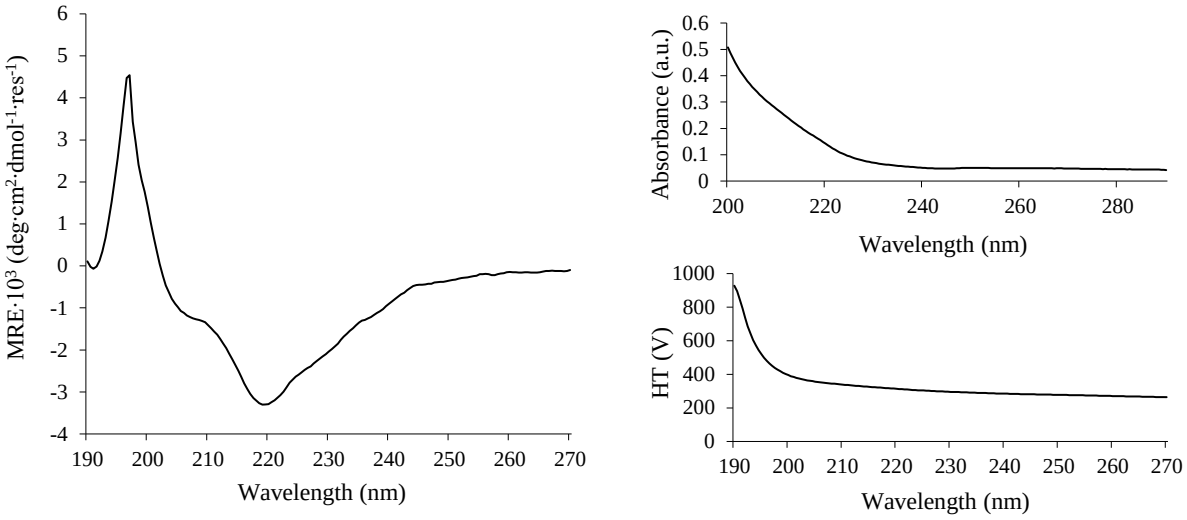


Figure 8.43: 41 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 7 µM peptide concentration in MeOH.

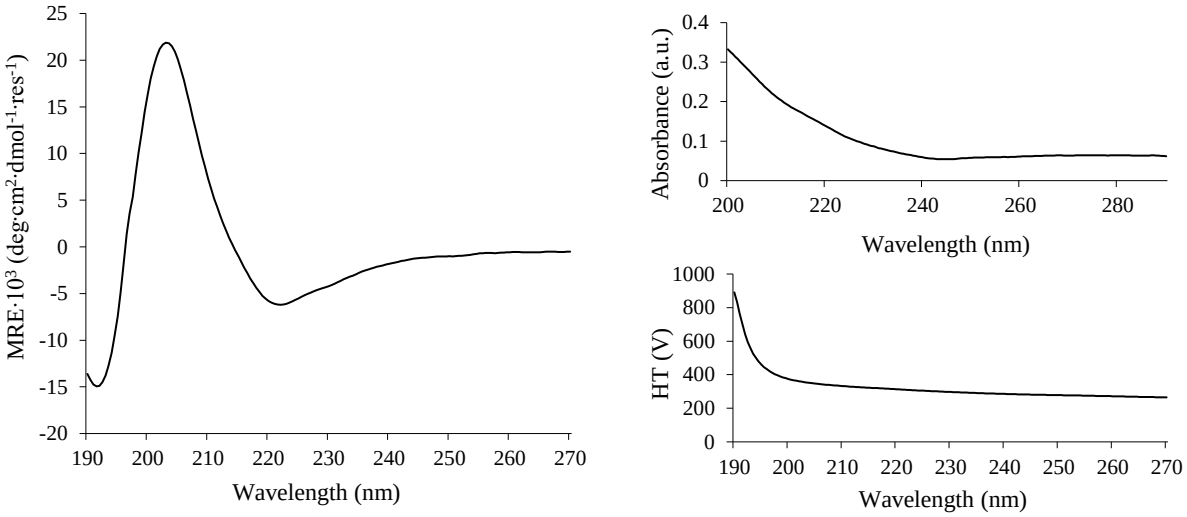


Figure 8.44: 42 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 5 µM peptide concentration in MeOH.

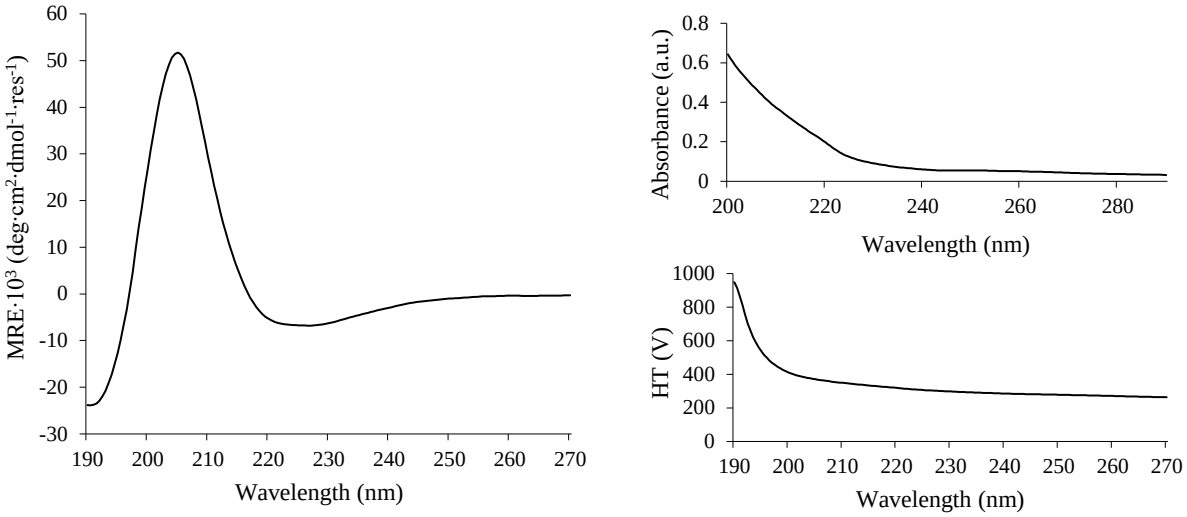


Figure 8.45: 43 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 9 µM peptide concentration in MeOH.

8.2.2. Chapter 4[xiv]

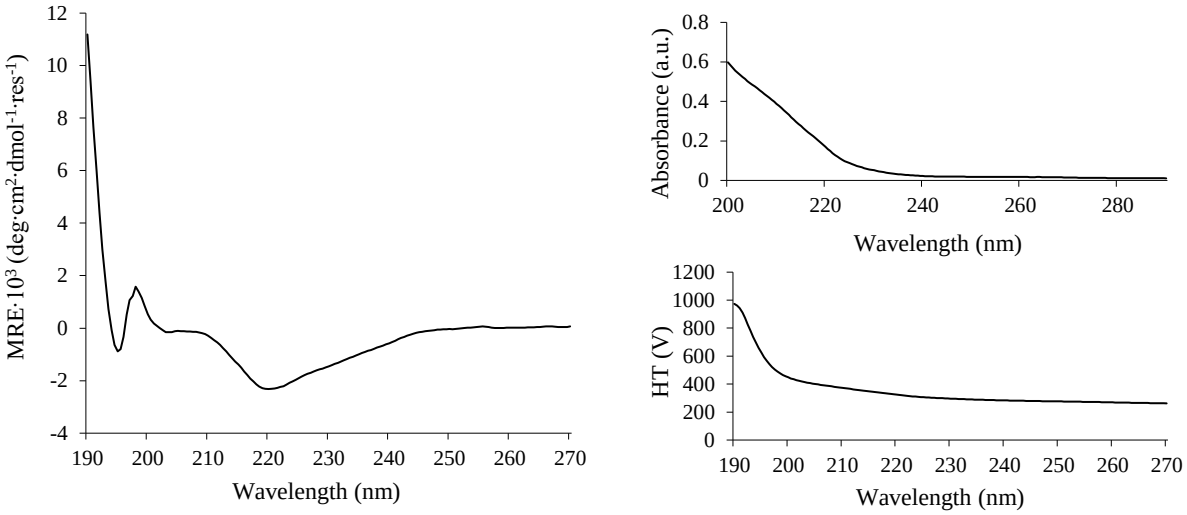


Figure 8.46: 45 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 125 µM peptide concentration in MeOH.

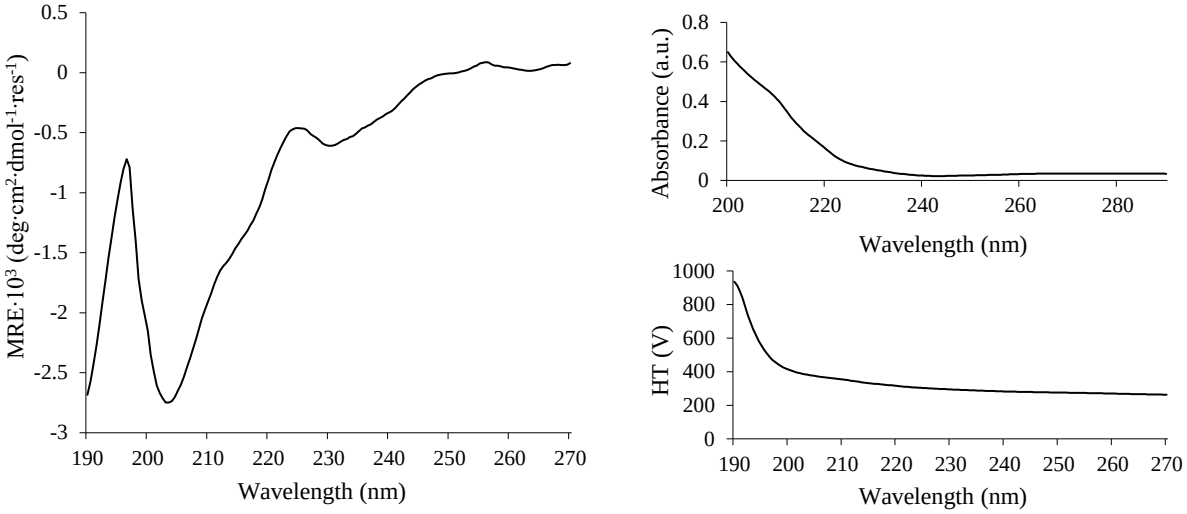


Figure 8.47: 46 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 125 µM peptide concentration in MeOH.

[xiv] CD spectra for **45** and **46r** was acquired by Danny Burke

8.2.3. Chapter 5

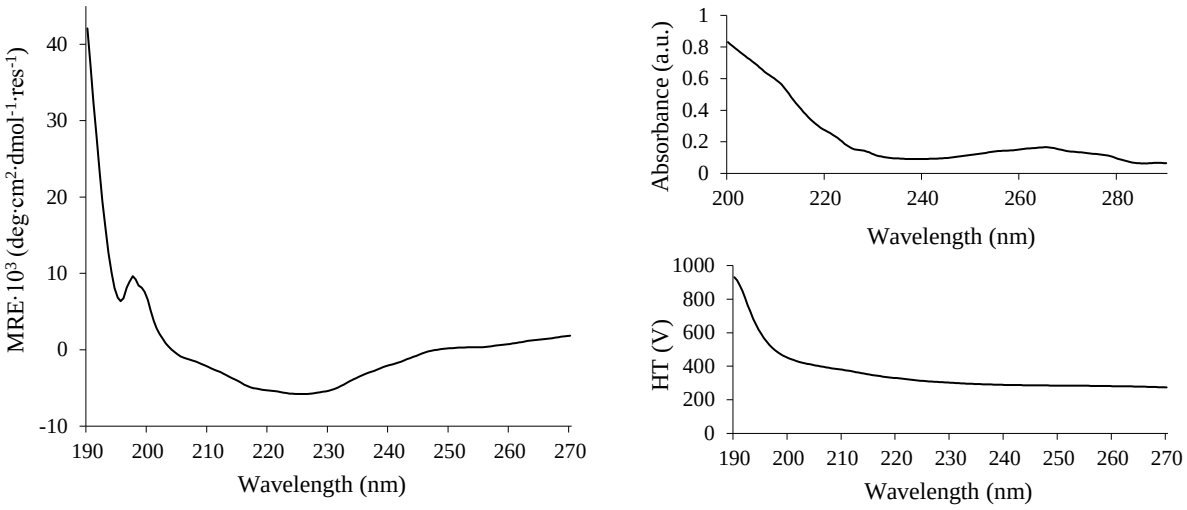


Figure 8.48: 54 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 95 µM peptide concentration in MeOH.

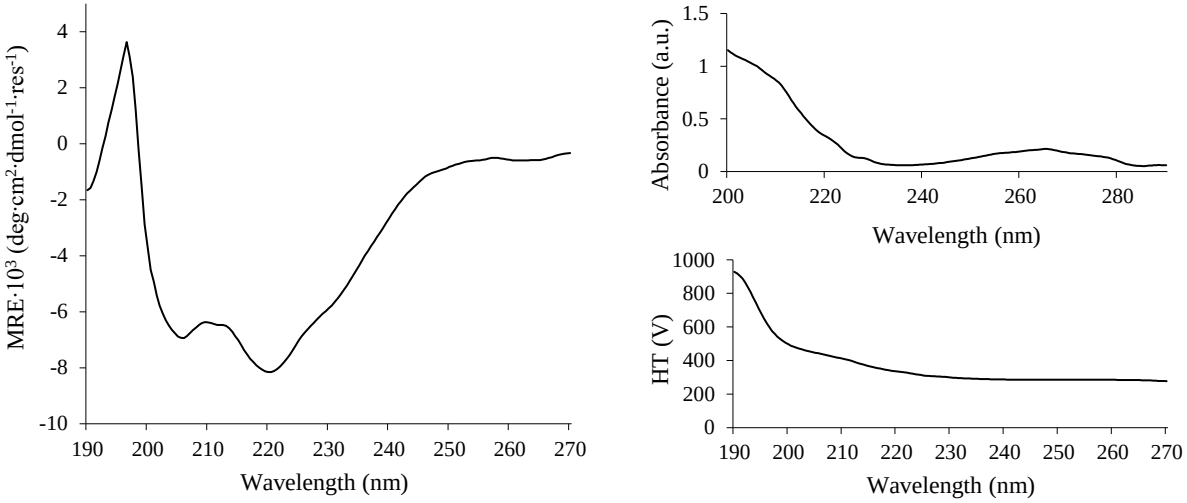


Figure 8.49: 55 - (left) CD spectrum at 25 °C. (right, top) Plot of high-tension voltage applied to the detector and plot of the absorbance (right, bottom). Conditions: 108 µM peptide concentration in MeOH.

8.3. ^{1}H, ROESY and NOESY NMR spectra of α/β-peptide decamers

34

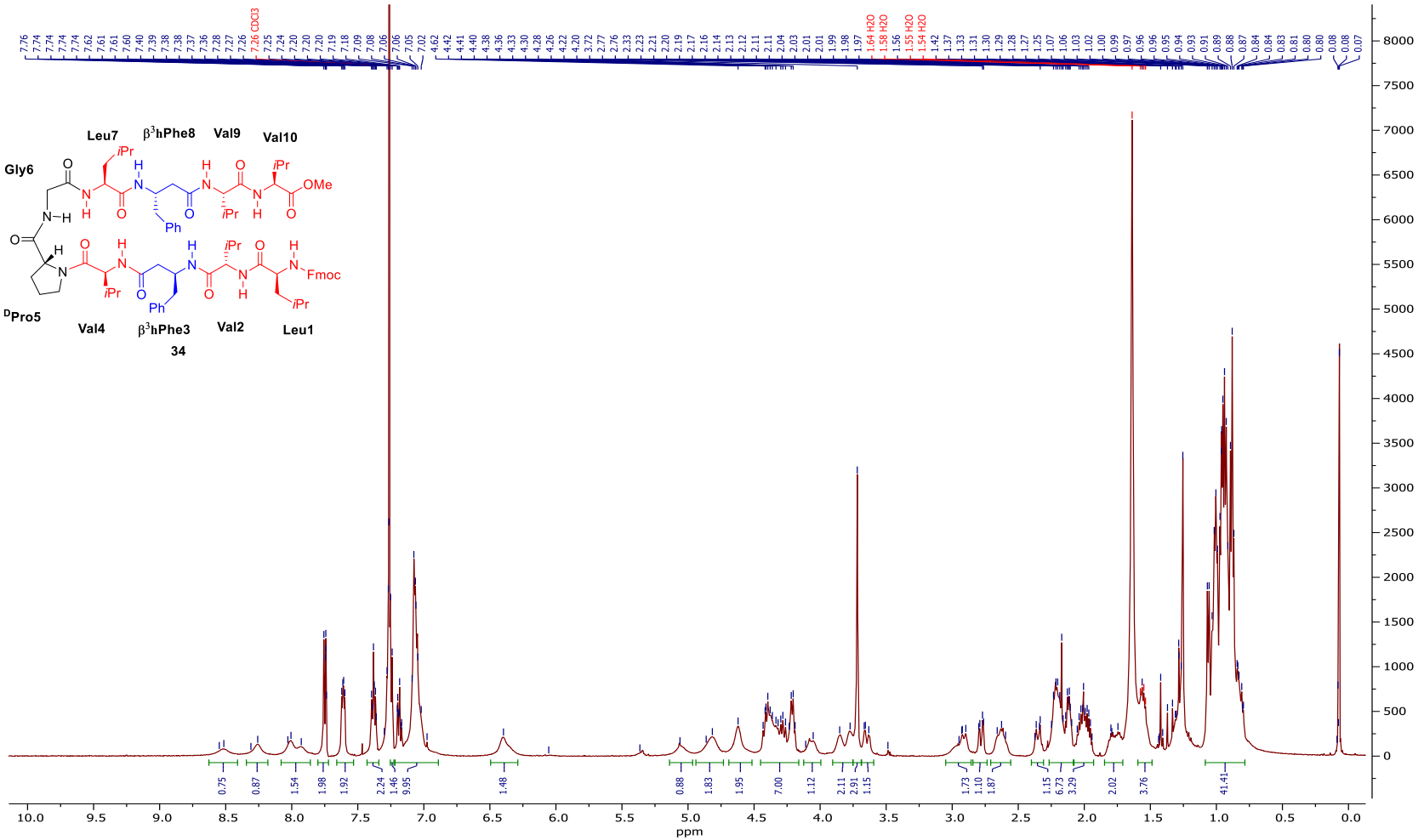


Figure 8.50: 500 MHz ^{1}H NMR spectrum of **34** (~4 mM) in $CDCl_3$.

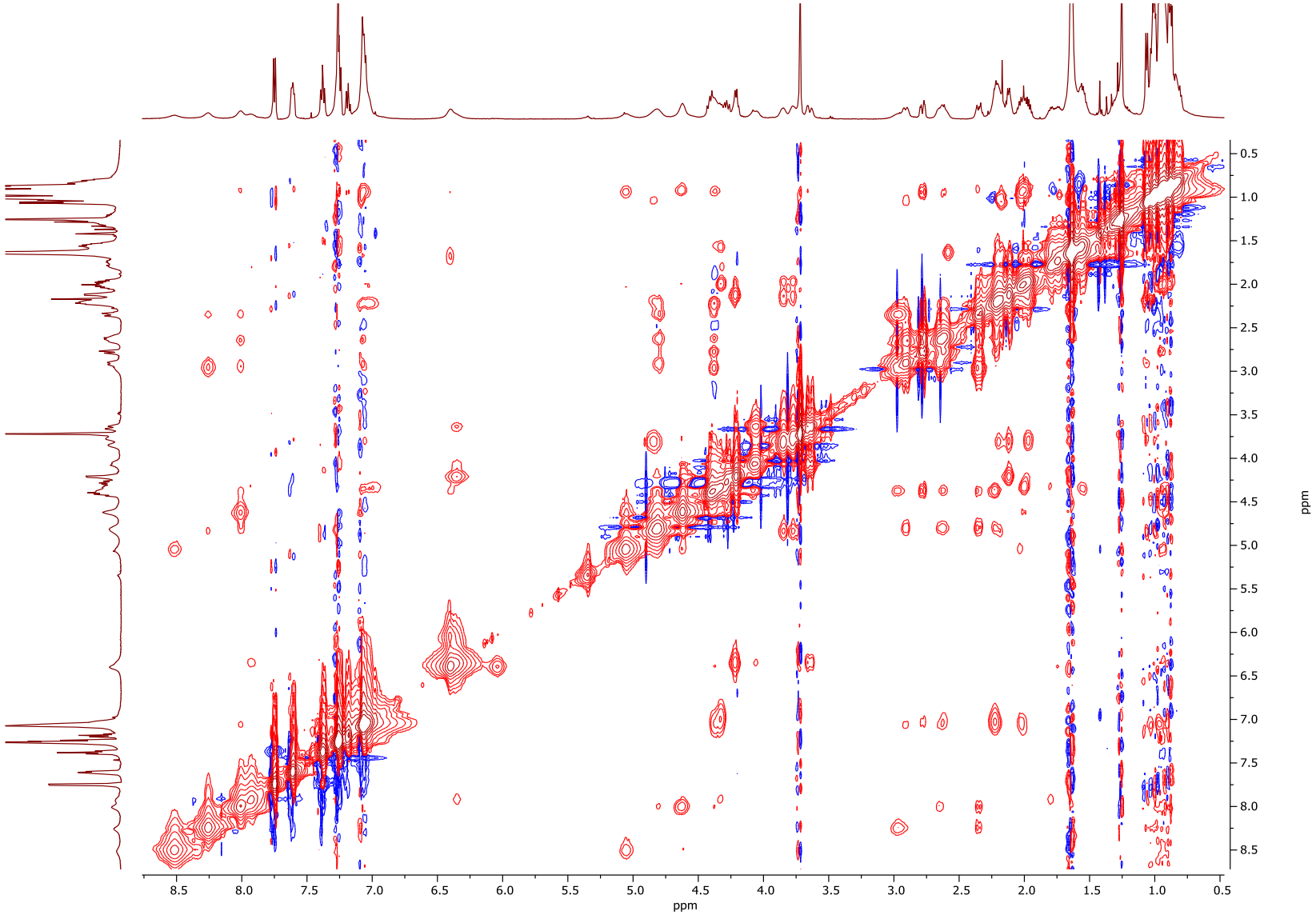


Figure 8.51: NOESYNMR spectrum of **34** (~4 mM) in $CDCl_3$.

35

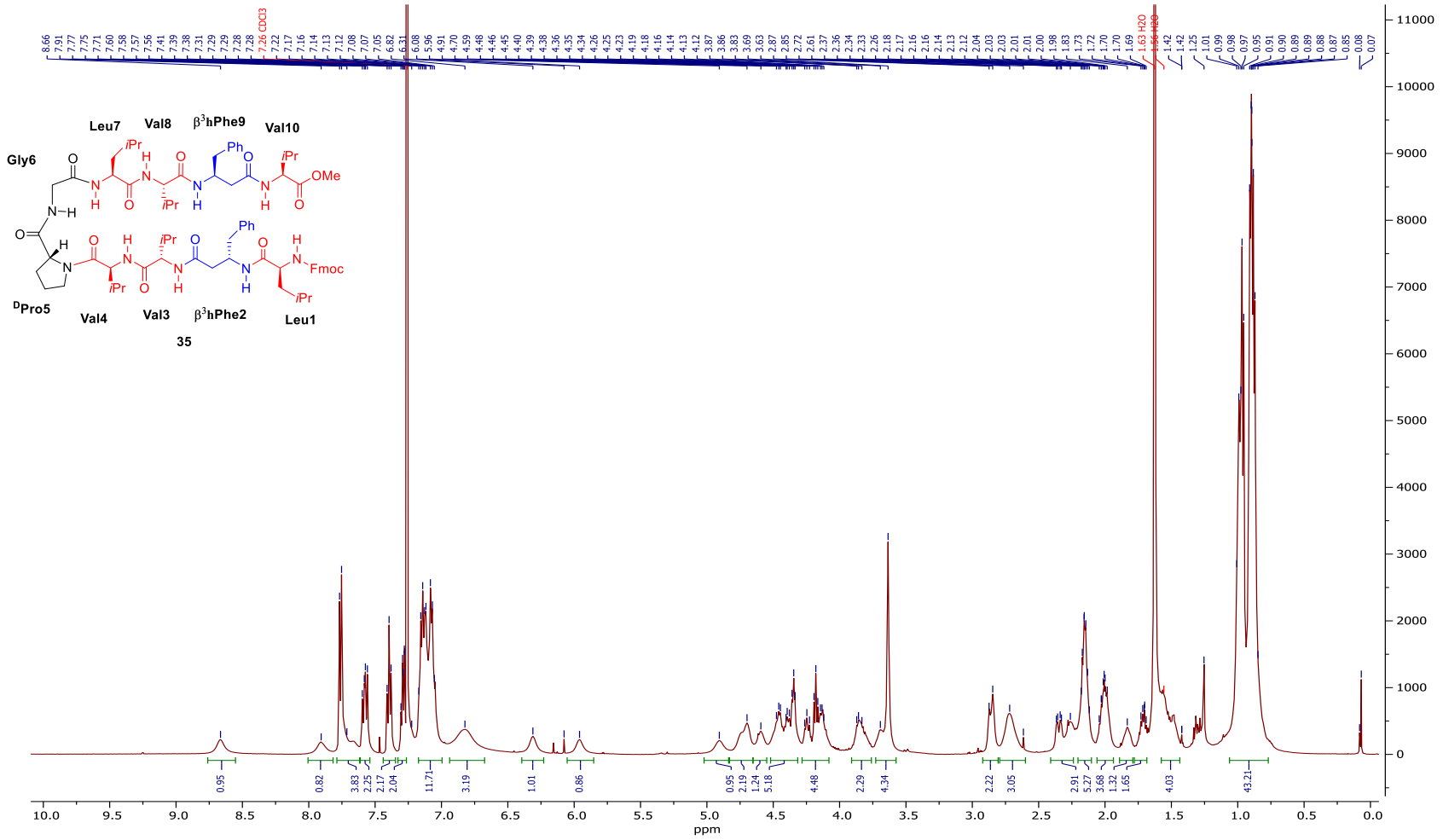


Figure 8.52: 500 MHz ^{1}H NMR spectrum of **35** (~4 mM) in $CDCl_3$.

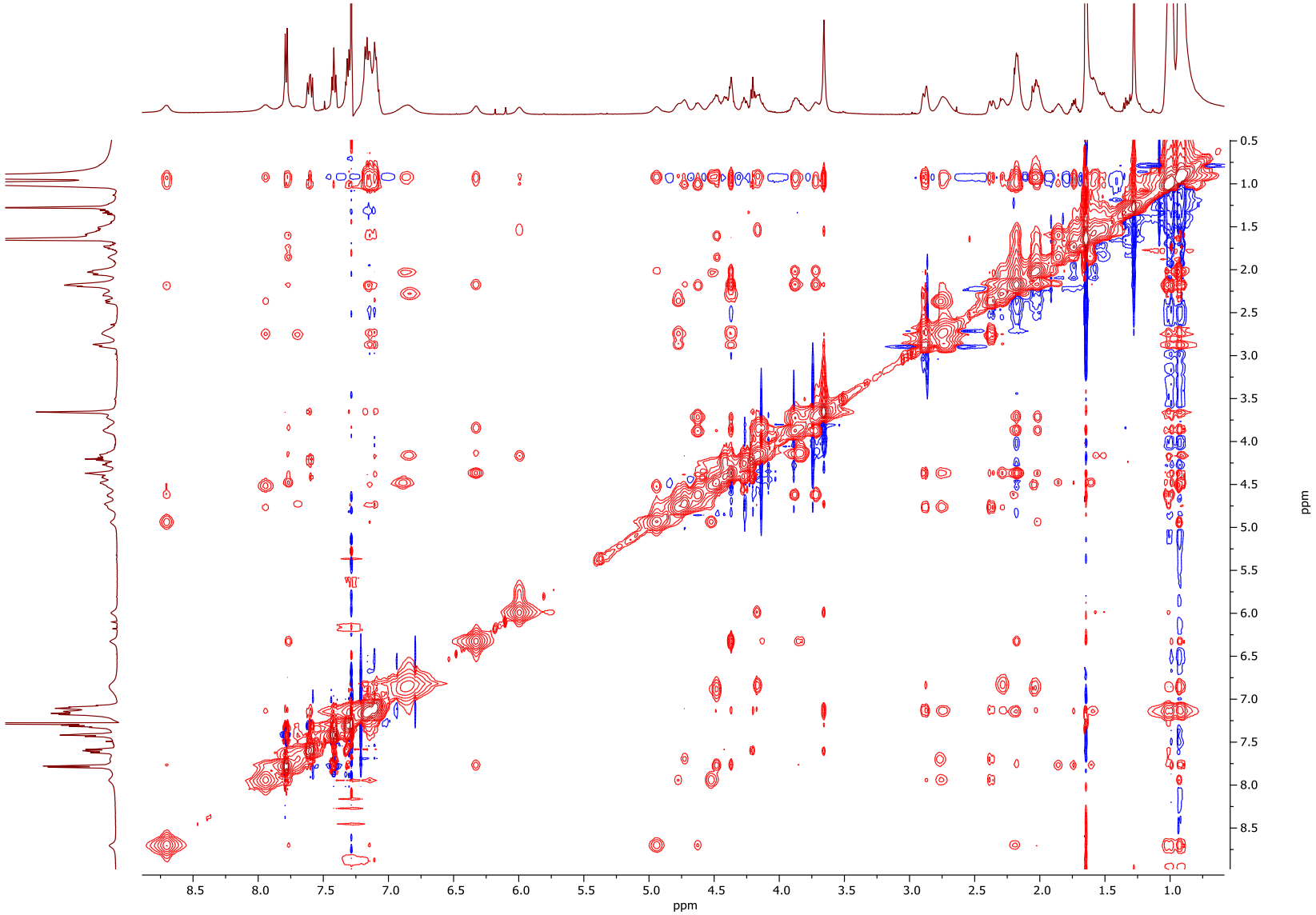


Figure 8.53: NOESYNMR spectrum of **35** (~4 mM) in $CDCl_3$.

36

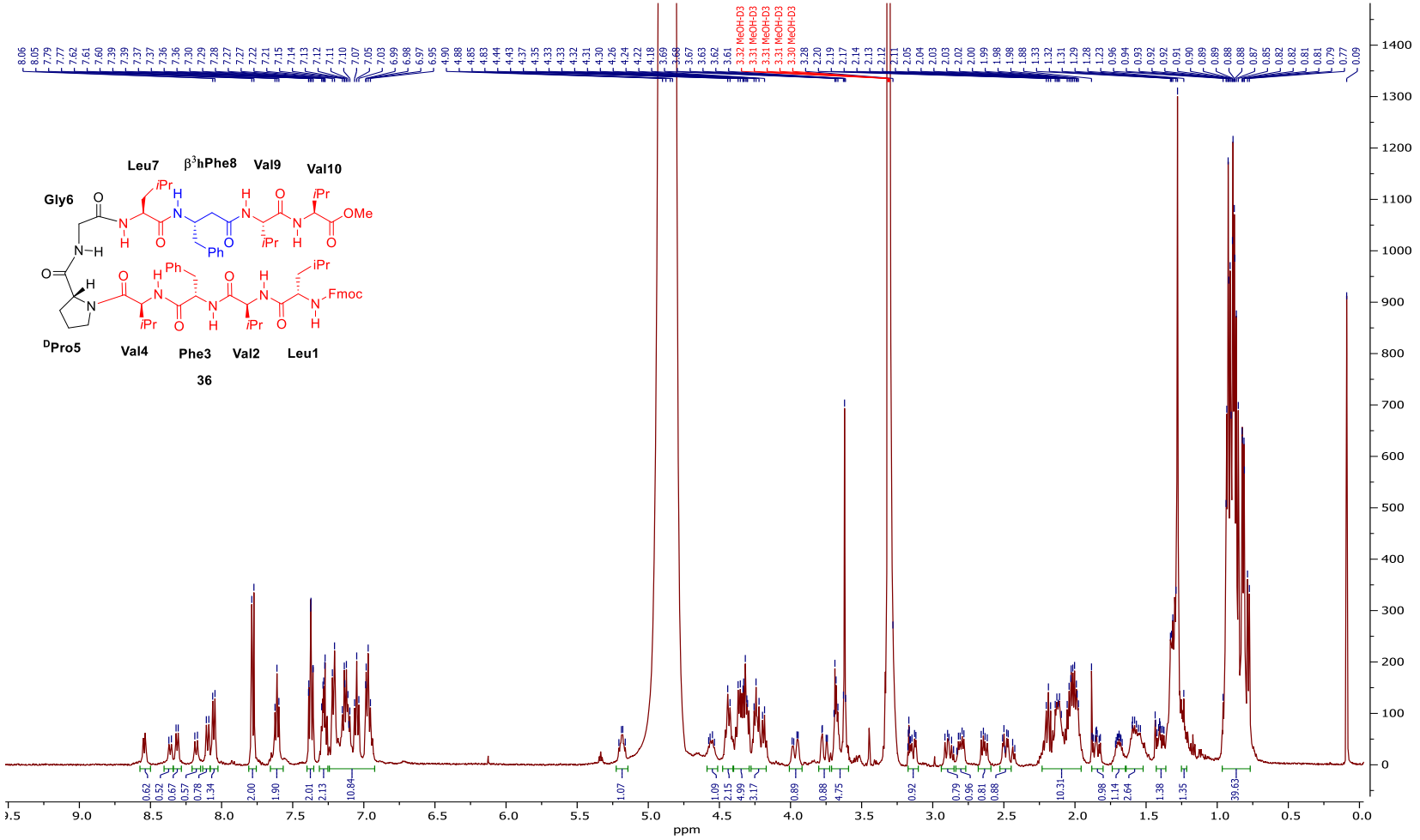


Figure 8.54: 500 MHz ^{1}H NMR spectrum of **36** (~1 mM) in CD_3OH.

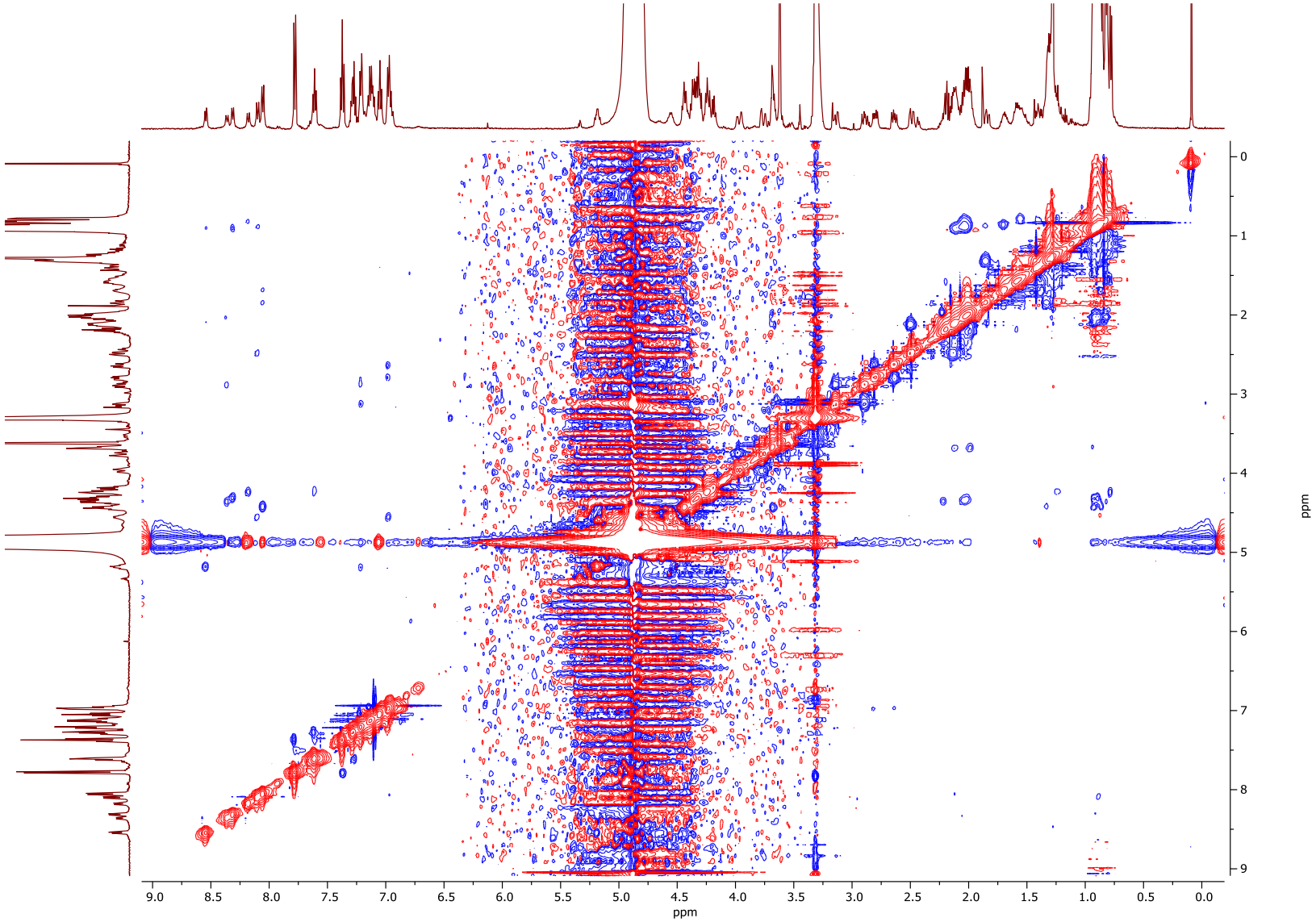


Figure 8.55: ROESY NMR spectrum of **36** (~1 mM) in CD_3OH.

38

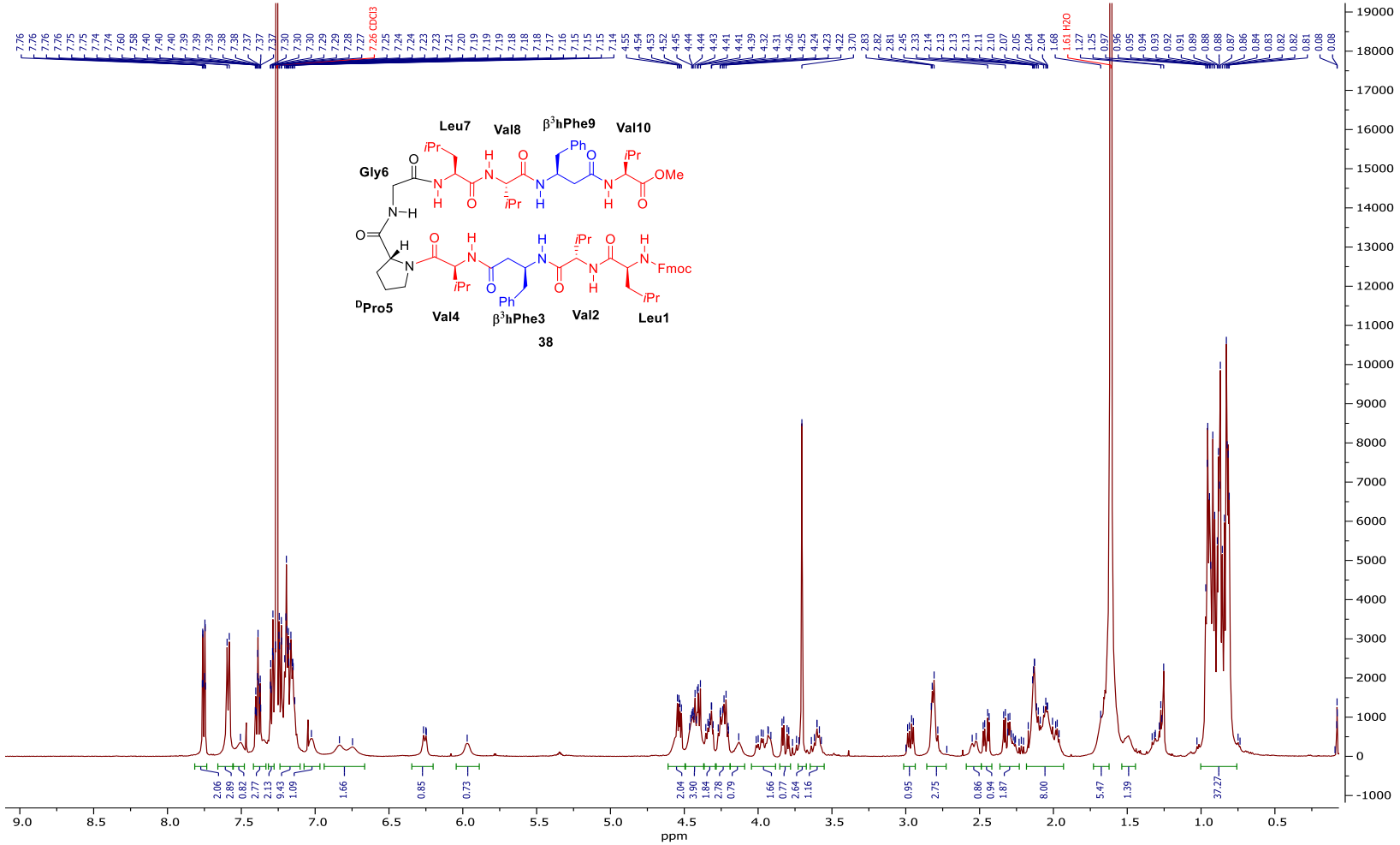


Figure 8.56: 500 MHz ^{1}H NMR spectrum of **38** (~2 mM) in $CDCl_3$.

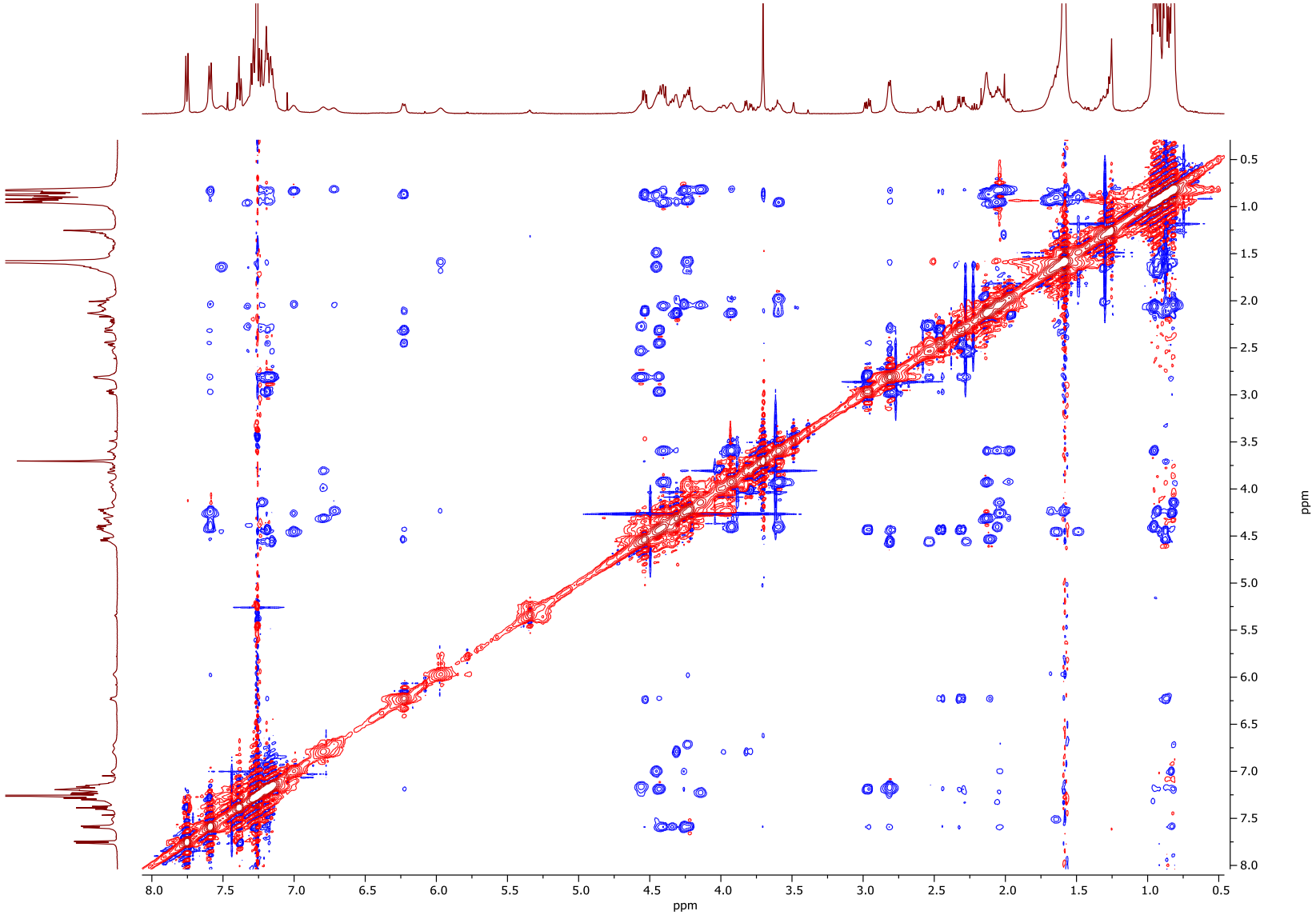


Figure 8.57: ROESY NMR spectrum of **38** (~2 mM) in $CDCl_3$.

8.4. Variable temperature studies

35

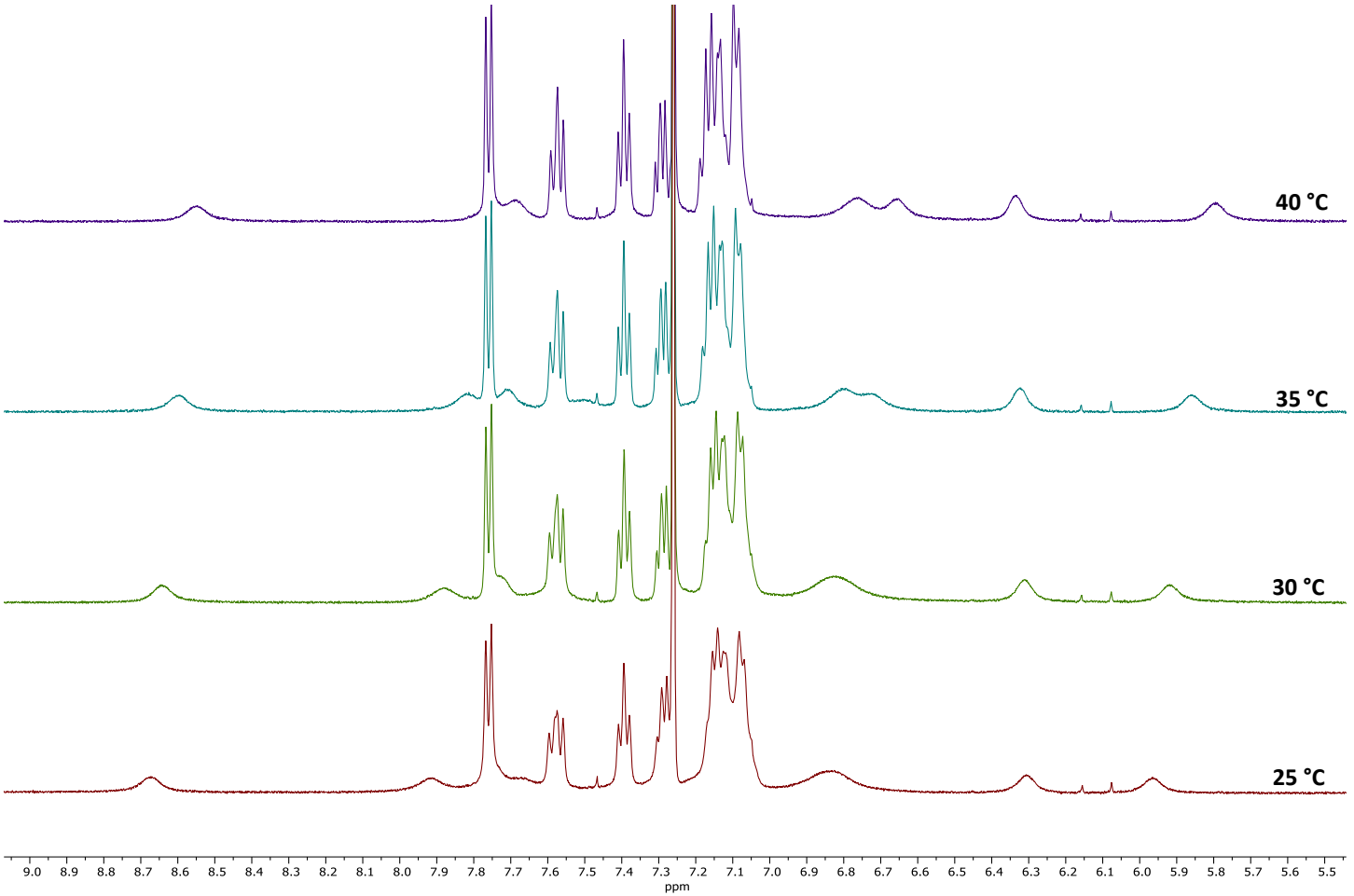


Figure 8.58: Partial 500 MHz ^{1}H NMR spectrum of **35** at various temperatures in $CDCl_3$.

36

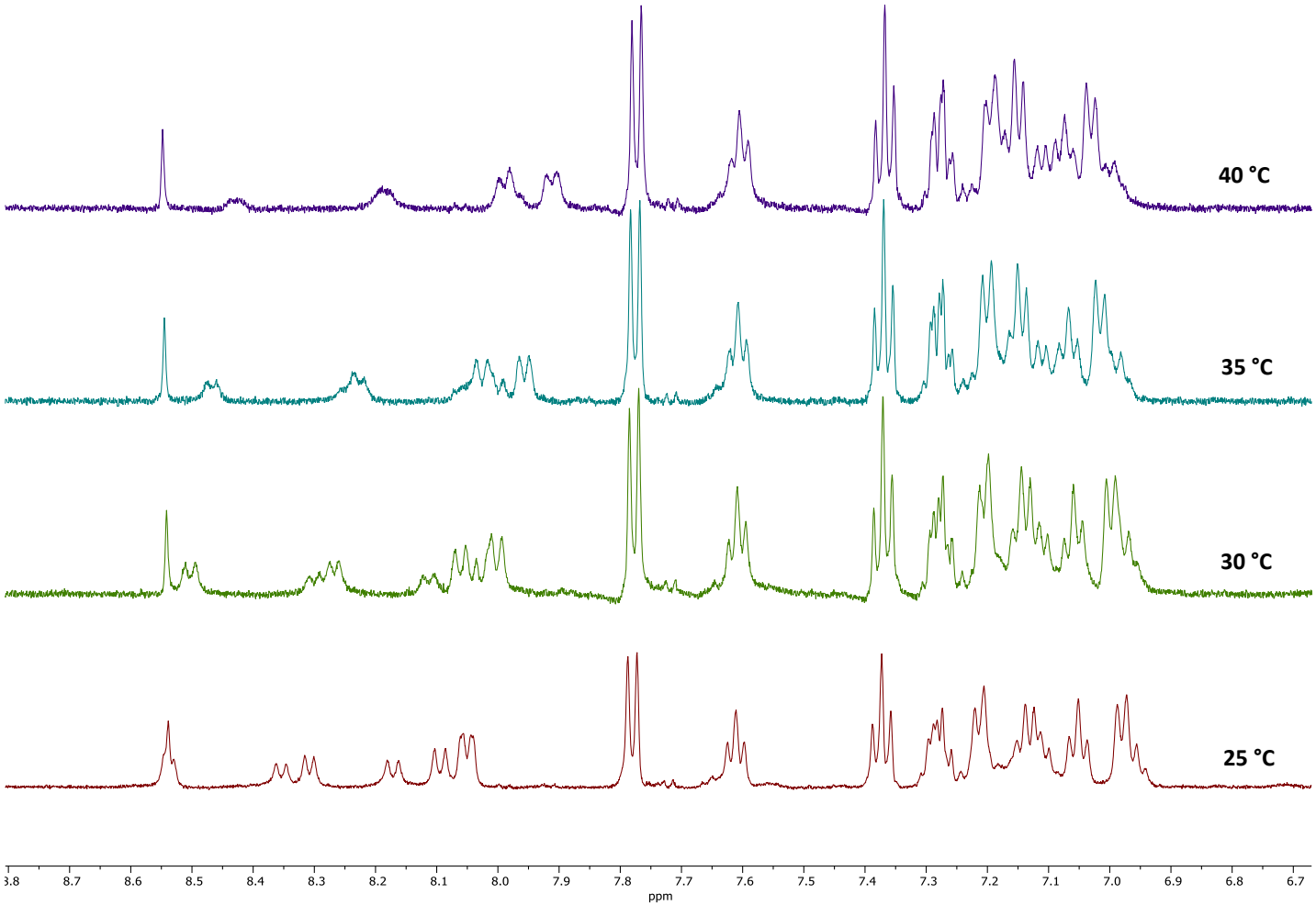


Figure 8.59: Partial 500 MHz ^{1}H NMR spectrum of **36** at various temperatures in CD_3OH.

References

(1) Cronin, L.; Walker, S. I. *Science* **2016**, *352*, 1174–1175.

(2) Aydinoglu, A. U.; Taşkın, Z. *Orig. Life Evol. Biosph.* **2018**, *48*, 55–71.

(3) Walker, S. I. *Nature* **2019**, *569*, 36–38.

(4) Mann, S. *Angew. Chem. Int. Ed.* **2013**, *52*, 155–162.

(5) Brewer, A.; Davis, A. P. *Nat. Chem.* **2013**, *6*, 569–574.

(6) Sutherland, J. D. *Angew. Chem. Int. Ed.* **2016**, *55*, 104–121.

(7) Bada, J. L. *Chem. Soc. Rev.* **2013**, *42*, 2186.

(8) Powner, M. W.; Sutherland, J. D. *Philos. Trans. R. Soc. B Biol. Sci.* **2011**, *366*, 2870–2877.

(9) Fiore, M.; Strazewski, P. *Angew. Chem. Int. Ed.* **2016**, *55*, 13930–13933.

(10) Weiss, M. C.; Sousa, F. L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W. F. *Nat. Microbiol.* **2016**, *1*, 16116.

(11) Cantine, M. D.; Fournier, G. P. *Orig. Life Evol. Biosph.* **2018**, *48*, 35–54.

(12) Woese, C. R.; Fox, G. E. *Proc. Natl. Acad. Sci.* **1977**, *74*, 5088–5090.

(13) Ruiz-Mirazo, K.; Briones, C.; De La Escosura, A. *Chem. Rev.* **2014**, *114*, 285–366.

(14) Krishnamurthy, R. *Acc. Chem. Res.* **2017**, *50*, 455–459.

(15) Joyce, G. F. *Angew. Chem. Int. Ed.* **2007**, *46*, 6420–6436.

(16) Mann, S. *Angew. Chemie Int. Ed.* **2008**, *47*, 5306–5320.

(17) Loakes, D.; Holliger, P. *Mol. Biosyst.* **2009**, *5*, 686.

(18) Joyce, G. F. *Nature* **2002**, *418*, 214–221.

(19) Smith, J. E.; Mowles, A. K.; Mehta, A. K.; Lynn, D. G. *Life* **2014**, *4*, 887–902.

(20) Islam, S.; Powner, M. W. *Chem* **2017**, *2*, 470–501.

(21) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920–930.

(22) Lincoln, T. A.; Joyce, G. F. *Science* **2009**, *323*, 1229–1232.

(23) Meyer, A. J.; Ellefson, J. W.; Ellington, A. D. *Acc. Chem. Res.* **2012**, *45*, 2097–2105.

(24) Horning, D. P.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **2016**, *113*, 9786–9791.

(25) Higgs, P. G.; Lehman, N. *Nat. Rev. Genet.* **2015**, *16*, 7–17.

(26) Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **2020**, *Early edit*.

(27) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459*, 239–242.

(28) Islam, S.; Bučar, D.-K.; Powner, M. W. *Nat. Chem.* **2017**, *9*, 584–589.

(29) Becker, S.; Schneider, C.; Dejmek, M.; Carell, T.; Okamura, H.; Crisp, A. *Nat. Commun.* **2018**, *9*, 163.

(30) Nam, I.; Nam, H. G.; Zare, R. N. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 36–40.

(31) Xu, J.; Tsanakopoulou, M.; Magnani, C. J.; Szabla, R.; Šponer, J. E.; Šponer, J.; Góra, R. W.; Sutherland, J. D. *Nat. Chem.* **2016**, *9*, 303–309.

(32) Sievers, D.; von Kiedrowski, G. *Nature* **1994**, *369*, 221–224.

(33) Giurgiu, C.; Li, L.; O'Flaherty, D. K.; Tam, C. P.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, *139*, 16741–16747.

(34) Jia, T. Z.; Fahrenbach, A. C.; Kamat, N. P.; Adamala, K. P.; Szostak, J. W. *Nat. Chem.* **2016**, *8*, 915–921.

(35) Engelhart, A. E.; Powner, M. W.; Szostak, J. W. *Nat. Chem.* **2013**, *5*, 390–394.

(36) He, C.; Gállego, I.; Laughlin, B.; Grover, M. A.; Hud, N. V. *Nat. Chem.* **2016**, *9*, 318–324.

(37) Cafferty, B. J.; Fialho, D. M.; Khanam, J.; Krishnamurthy, R.; Hud, N. V. *Nat. Commun.* **2016**, *7*, 11328.

(38) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183–187.

(39) Roberts, S. J.; Szabla, R.; Todd, Z. R.; Stairs, S.; Bučar, D. K.; Šponer, J.; Sasselov, D. D.; Powner, M. W. *Nat. Commun.* **2018**, *9*, 1–10.

(40) Anastasi, C.; Buchet, F. F.; Crowe, M. A.; Parkes, A. L.; Powner, M. W.; Smith, J. M.; Sutherland, J. D. *Chem. Biodivers.* **2007**, *4*, 721–739.

(41) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. *Nature* **1996**, *382*, 525–528.

(42) Yao, S.; Ghosh, I.; Zutshi, R.; Chmielewski, J. *J. Am. Chem. Soc.* **1997**, *119*, 10559–10560.

(43) Issac, R.; Ham, Y.-W.; Chmielewski, J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 458–463.

(44) Bissette, A. J.; Fletcher, S. P. *Angew. Chem. Int. Ed.* **2013**, *52*, 12800–12826.

(45) Severin, K.; Lee, D. H.; Martinez, J. A.; Vieth, M.; Ghadiri, M. R. *Angew. Chem. Int. Ed.* **1998**, *37*, 126–128.

(46) Saghatellan, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. *Nature* **2001**, *409*, 797–801.

(47) Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. *Nature* **1998**, *390*, 591–594.

(48) Rubinov, B.; Wagner, N.; Rapaport, H.; Ashkenasy, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 6683–6686.

(49) Rubinov, B.; Wagner, N.; Matmor, M.; Regev, O.; Ashkenasy, N.; Ashkenasy, G. *ACS Nano* **2012**, *6*, 7893–7901.

(50) Nanda, J.; Rubinov, B.; Ivnitski, D.; Mukherjee, R.; Shtelman, E.; Motro, Y.; Miller, Y.; Wagner, N.; Cohen-Luria, R.; Ashkenasy, G. *Nat. Commun.* **2017**, *8*, 434.

(51) Rout, S. K.; Friedmann, M. P.; Riek, R.; Greenwald, J. *Nat. Commun.* **2018**, *9*, 234.

(52) Carnall, J. M. A.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C. A.; Peyralans, J. J.-P.; Otto, S. *Science* **2010**, *327*, 1502–1506.

(53) Malakoutikhah, M.; Peyralans, J. J.-P.; Colomb-Delsuc, M.; Fanlo-Virgós, H.; Stuart, M. C. A.; Otto, S. *J. Am. Chem. Soc.* **2013**, *135*, 18406–18417.

(54) Sadownik, J. W.; Mattia, E.; Nowak, P.; Otto, S. *Nat. Chem.* **2016**, *8*, 264–269.

(55) Goodwin, J. T.; Mehta, A. K.; Lynn, D. G. *Acc. Chem. Res.* **2012**, *45*, 2189–2199.

(56) Maury, C. P. J. *Orig. Life Evol. Biosph.* **2009**, *39*, 141–150.

(57) Danger, G.; Plasson, R.; Pascal, R. *Chem. Soc. Rev.* **2012**, *41*, 5416–5429.

(58) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Science* **2008**, *322*, 404–404.

(59) Amend, J. P.; Shock, L. E. *Science* **1998**, *281*, 1659–1662.

(60) Longo, L. M.; Blaber, M. *Front. Microbiol.* **2013**, *4*, 2013–2015.

(61) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. *Nat. Chem.* **2015**, *7*, 301–307.

(62) Burton, A. S.; Stern, J. C.; Elsila, J. E.; Glavin, D. P.; Dworkin, J. P. *Chem. Soc. Rev.* **2012**, *41*, 5459–5472.

(63) Gibard, C.; Bhowmik, S.; Karki, M.; Kim, E. K.; Krishnamurthy, R. *Nat. Chem.* **2018**, *10*, 212–217.

(64) Parker, E. T.; Zhou, M.; Burton, A. S.; Glavin, D. P.; Dworkin, J. P.; Krishnamurthy, R.; Fernández, F. M.; Bada, J. L. *Angew. Chem. Int. Ed.* **2014**, *53*, 8132–8136.

(65) Greenwald, J.; Friedmann, M. P.; Riek, R. *Angew. Chem. Int. Ed.* **2016**, *55*, 11609–11613.

(66) Campbell, T. D.; Febrian, R.; McCarthy, J. T.; Kleinschmidt, H. E.; Forsythe, J. G.; Bracher, P. J. *Nat. Commun.* **2019**, *10*, 4508.

(67) Smith, A. J.; Ali, F. I.; Soldatov, D. V. *CrystEngComm* **2014**, *16*, 7196–7208.

(68) Cheng, R. P.; Gellman, S. H.; Degrado, W. F. *Chem. Rev.* **2001**, *101*, 3219–3232.

(69) Seebach, D.; Beck, A. K.; Bierbaum, D. J. *Chem. Biodivers.* **2004**, *1*, 1111–1239.

(70) Robinson, J. a. *Acc. Chem. Res.* **2008**, *41*, 1278–1288.

(71) Robinson, J. A. *Synlett* **1999**, *4*, 429–441.

(72) Gopi, H. N.; Roy, R. S.; Raghothama, S. R.; Karle, I. L.; Balaram, P. *Helv. Chim. Acta* **2002**, *85*, 3313–3330.

(73) Huck, B. R.; Fisk, J. D.; Gellman, S. H. *Org. Lett.* **2000**, *2*, 2607–2610.

(74) Krauthauser, S.; Christianson, L. A.; Popwel, D. R.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 11719–11720.

(75) Karle, I. L.; Gopi, H. N.; Balaram, P. *Proc. Natl. Acad. Sci.* **2001**, *98*, 3716–3719.

(76) Roy, R. S.; Gopi, H. N.; Raghothama, S.; Gilardi, R. D.; Karle, I. L.; Balaram, P. *Biopolym. - Pept. Sci. Sect.* **2005**, *80*, 787–799.

(77) Sonti, R.; Gopi, H. N.; Muddegowda, U.; Ragothama, S.; Balaram, P. *Chem. - A Eur. J.* **2013**, *19*, 5955–5965.

(78) Segman, S.; Lee, M.-R.; Vaiser, V.; Gellman, S. H.; Rapaport, H. *Angew. Chem. Int. Ed.* **2010**, *49*, 716–719.

(79) Schmitt, M. A.; Choi, S. H.; Guzei, I. A.; Gellman, S. H. *J. Am. Chem. Soc.* **2005**, *127*, 13130–13131.

(80) Segman-Magidovich, S.; Lee, M. R.; Vaiser, V.; Struth, B.; Gellman, S. H.; Rapaport, H. *Chem. - A Eur. J.* **2011**, *17*, 14857–14866.

(81) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074.

(82) Cowper, B.; Sze, T. M.; Premdjee, B.; Bongat White, A. F.; Hacking, A.; Macmillan, D. *Chem. Commun.* **2015**, *51*, 3208–3210.

(83) Kajihara, Y.; Yoshihara, A.; Hirano, K.; Yamamoto, N. *Carbohydr. Res.* **2006**, *341*, 1333–1340.

(84) Cowper, B.; Sze, T. M.; Premdjee, B.; Bongat White, A. F.; Hacking, A.; Macmillan, D. *Chem. Commun.* **2015**, *51*, 3208–3210.

(85) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.

(86) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, *119*, 4325–4329.

(87) Kemp, D. S.; Rebek, J. J. *J. Am. Chem. Soc.* **1970**, *595*, 5792–5793.

(88) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10068–10073.

(89) Karle, I. L.; Awasthi, S. K.; Balaram, P. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8189–8193.

(90) Awasthi, S. K.; Raghothama, S.; Balaram, P. *Biochem. Biophys. Res. Commun.* **1995**, *216*, 375–381.

(91) Karle, I.; Gopi, H. N.; Balaram, P. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5160–5164.

(92) Sonti, R.; Gopi, H. N.; Muddegowda, U.; Ragothama, S.; Balaram, P. *Chem. - A Eur. J.* **2013**, *19*, 5955–5965.

(93) Williamson, M. P.; Waltho, J. P. *Chem. Soc. Rev.* **1992**, *21*, 227–236.

(94) Moretto, V.; Crisma, M.; Bonora, G. M.; Toniolo, C.; Balaram, P.; Balaram, H. *Macromolecules* **1989**, *22*, 2939–2944.

(95) Tomlinson, J. H.; Williamson, M. P. *J Biomol NMR* **2012**, *52*, 57–64.

(96) Roy, R. S.; Gopi, H. N.; Raghothama, S.; Karle, I. L.; Balaram, P. *Chem. - A Eur. J.* **2006**, *12*, 3295–3302.

(97) Hegedüs, Z.; Wéber, E.; Kriston-Pál, É.; Makra, I.; Czibula, Á.; Monostori, É.; Martinek, T. A. *J. Am. Chem. Soc.* **2013**, *135*, 16578–16584.

(98) Kung, V. M.; Cornilescu, G.; Gellman, S. H. *Angew. Chem. Int. Ed.* **2015**, *54*, 14336–14339.

(99) Nesloney, C. L.; Kelly, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 5836–5845.

(100) Brahms, S.; Brahms, J.; Spach, G.; Brack, A. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 3208–3212.

(101) Bush, C. A.; Sarkar, S. K.; Kopple, K. D. *Biochemistry* **1978**, *17*, 4951–4954.

(102) Freire, F.; Almeida, A. M.; Fisk, J. D.; Steinkruger, J. D.; Gellman, S. H. *Angew. Chemie* **2011**, *123*, 8894–8897.

(103) Hamley, I. W. *Biomacromolecules* **2014**, *15*, 1543–1559.

(104) Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Adv. Drug Deliv. Rev.* **2002**, *64*, 459–457.

(105) Segman, S.; Lee, M. R.; Vaiser, V.; Gellman, S. H.; Rapaport, H. *Angew. Chem. Int. Ed.* **2010**, *49*, 716–719.

(106) Aapptec. Practical Synthesis Guide to Solid Phase Peptide Chemistry http://www.aapptec.com/solid-phase-peptide-synthesis-i-241.html.

(107) Ueda, T.; Sada, I.; Kato, T.; Izumiya, N. *Int. J. Pept. Protein Res.* **1985**, *25*, 475–480.

(108) Bachem Peptide Guide http://www.bachem.com/ (accessed Jul 1, 2016).

(109) Yan, B.; Zhang, B. *Analytical Methods in Combinatorial Chemistry*; Critical Reviews in Combinatorial Chemistry; CRC Press, 2010.

(110) Kuipers, B. J. H.; Gruppen, H. *J. Agric. Food Chem.* **2007**, *55*, 5445–5451.